ECO-EPIDEMIOLOGY OF WEST NILE VIRUS IN THE MIDWESTERN UNITED STATES

BY

JOHNNY ALBERT UELMEN JR

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in VMS-Pathobiology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Doctoral Committee:

Associate Professor Rebecca Smith, Chair
Assistant Professor Keith Jarosinski
Associate Professor Brian Allan
Associate Professor Jennifer Fraterrigo
Associate Professor Sadie Ryan, University of Florida - Gainesville
Professor Jonathan Patz, University of Wisconsin - Madison

Clinical Associate Professor Marilyn O'Hara Ruiz (deceased)

ABSTRACT

Infectious diseases are on the rise globally. Although only accounting for 17% of all infectious diseases, vector-borne diseases are increasing the fastest. Global changes in climate, particularly in precipitation and temperature, directly affect the life cycle of arthropod vectors, including the incubation period for any pathogen they transmit. Mosquitoes are among the most responsive to climatic changes, and several species are considered the most successful invasive organisms on our planet. In 1999, West Nile virus (WNV) arrived in New York City and caused a local outbreak affecting birds and humans. In four years, the virus reached California and was present in nearly every state. Two decades later, WNV is now the most important mosquito-borne disease in North America and continues to cause human infection and death. Countless resources, person-hours, and millions of dollars have been used to control and monitor WNV throughout the country. However, predicting when and where infection will occur has proven immensely difficult, largely due to the complex relationships of numerous interacting factors affecting WNV disease ecology. Considerable variation in factors affecting WNV transmission derive from climatic, environmental, physical, and human socio-economic and demographic forces, and can vary weekly and across fine-scales. Several research teams have investigated these dynamics, attempting to improve our understanding of the drivers of WNV. While research is improving many aspects of mosquito control and mitigating risk, human infection continues. Additionally, results and interpretations of WNV are less clear, sometimes conflicting among overlapping study areas. The state of Illinois, experiencing the first positive mosquitoes in 2001, has since produced the fifth

most human cases in the country. Fortunately, the Smith and O'Hara Ruiz labs at the University of Illinois have dedicated over 10 years of research in improving our understanding of transmission in the Chicago region. Several key findings generated from these labs include the link between housing age and infection, key associations derived from complex spatial epidemiology methods, and human activity and behavior risk. The O'Hara Ruiz lab was also among the first research groups to associate strong relationships between mosquito infection with temporal lags in precipitation and temperature. Largely expanding from previous efforts in our lab, the aims of this dissertation are to create robust and accurate WNV forecast models for the midwestern United States. Specifically, the chapters of this dissertation focus on very fine-scale drivers of disease, and then comparatively analyze best-fit models across scales in Chicago. Lastly, this dissertation expands efforts from Chicago to the Midwest, evaluating mosquito infection across 118 counties and 8 states. Findings from this research demonstrated that drivers of WNV vary by scale. Specifically, the finer the scale, the more important an included covariate becomes. However, as scales become broader, the overall performance of WNV models increases. This research also found that precipitation and temperature (and their respective 2 and 3 week lags) are consistently the most important covariates in WNV transmission, corroborating several studies. However, acquiring additional data sources does improve model strength, but the overall net benefit given allocation of resources may not be the most efficient use of time and effort. Lastly, this study found that the upper Midwest, overall, has adequate spatial coverage of mosquito infection through current surveillance practices. However, there are several notable gaps in our understanding of disease, particularly in most of Iowa and southern

Wisconsin. Additionally, mosquito infection is increasing by about 14.2% annually, and is not an artifact of increased mosquito control efforts. Notable human outbreaks in 2005-2006, 2012, and 2018 coincide with years of highest mosquito infection in the Midwest. The time between these outbreaks is decreasing by about 1 year. The knowledge gained from this dissertation provide important, but sobering insights into the projections of human WNV infection in the Midwest. There is little doubt that future outbreaks will not only occur, but increase in frequency and numbers of infected. The models created and compared in these analyses can be used to understand trends and forecasts of WNV in other regions of the United States, as well as with other mosquito-borne diseases.

ACKNOWLEDGEMENTS

Words cannot express my heartfelt appreciation for Dr. Marilyn O'Hara Ruiz. Taken from us far too early, she was a leader in spatial epidemiology, and leaves behind a legacy that will forever be remembered. I am incredibly grateful for having the opportunity to be one of her students and learn directly from her during the first half of my Ph.D. studies at Illinois. Dr. O'Hara Ruiz was the sole reason I transferred from my doctoral program at the University of Wisconsin to the University of Illinois. Her expertise in the field and mutual interest in research brought me to Urbana, but it was her incredible personality, mentorship, and genuine care for her students and colleagues that taught my greatest lessons. I am blessed to honor her by carrying on her legacy.

Dr. Rebecca Smith demonstrated firsthand how to be a leader and calming force. I am so thankful she accepted me into her lab, essentially allowing me to maintain my duties and projects I had be working on with Dr. O'Hara Ruiz. As I became more advanced in spatial statistics, Dr. Smith trusted me and allowed my work and research ideas to flourish, herself expanding her background in spatial epidemiology and vector-borne disease ecology. Somehow, in light of losing such an amazing mentor, Dr. Smith helped make my research experience a beacon of hope and resilience. As a young assistant professor, Dr. Smith was always energetic, positive, and prioritized her students. Always open to not just research inquiries, but general questions about life, she showed me how to balance work and family. Now, as a newly and well deserved associate professor, Dr. Smith has shown me what it takes to be an accomplished researcher, phenomenal mentor,

and devoted parent. I am so excited to continue working with you as a post-doctoral research associate. Thank you for everything.

It only makes sense that the amazing mentors I've had at Illinois would likely recruit equally amazing students. I have had so much fun, countless jokes, and outbursts of laughter with my labmates. Thank you Lee Ann, Maya, Jameson, Sulagna, Bennett, Melanie, Alvyn, and Serena. You each provided a supportive and fun environment that reminded me to be me. I am very grateful to Bill Brown. The "Wizard" as I call him, because of his profound ability to get me whatever spatial data I needed super-fast. Bill, always so supportive and kind, never hesitated to help any of us. He helped me become a better user of GIS and understand the bigger picture. He likely won't take credit for any of this, but it's all true! I also want to give thanks to Dr. Surendra Karki, by which my motivation and overall research questions were based on. His previous work set the stage for my own. Surendra was an excellent student and later post-doc, and I am very thankful for having the opportunity to ask him so many questions relating to our disease system.

I want to express my appreciation for all those who worked with me, in the field, and remotely, to allow me to conduct my research. All those in Chicago, especially Dr. Patrick Irwin and Dan Bartlett, for supporting my research efforts in the Northwest Mosquito Abatement District. I've come into contact with so many other wonderful mosquito control people, and am grateful for all their help in data sharing and general conversations. In particular, I want to recognize Sam Force and Jason Probus at Macon Mosquito Control for all their support and kindness throughout all our interactions. I really appreciate you guys! As part of the Midwest Center of Excellent, I have had the

privilege to meet all these individuals, plus so many more. These opportunities have increased collaboration and provided me new and creative research ideas.

I would also like to thank my Department and Committee members for being supportive and flexible throughout the many changes and obstacles during my doctoral studies. Initially, I likely appeared as an oddity to the department and to many of those I worked with, as I do not necessarily fit into any one particular field of research. However, it was my goal to develop a dissertation that tied it all together and I am grateful to each of you for guiding me along the way.

Lastly, my driving force and passion is because of my family. Thank you to my parents, Mom, Dad, Barry, and Linda, siblings, Nikki, Julian, Molly, Gili, and Meghan, and aunts and uncles, especially Sue and Ed Brinson, for always supporting and believing in me. Most importantly, I am grateful for my wife, Lana. This past year we faced the biggest scare of our lives, but we came through the other side stronger.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1. BACKGROUND

In the age of the Anthropocene, our planet is experiencing unprecedented changes that are rapidly being shaped and influenced by the effects of human activity (Myers et al. 2013, Whitmee et al. 2015, The Lancet 2019). Human population has swelled from 1 billion before the 17th century to more than 7 billion at the turn of the 21st century (Bongaarts 2009). Technological advances, particularly in the manufacturing and agricultural industries, have provided our species with the resources to support our massive population (Betoret and Betoret 2020). However, these inherent successes do not come without a cost (Haines et al. 2019). Effects from human developments are unmistakably affecting the planet's natural ecosystems and ecological balances in a uniformly negative way (Whitmee et al. 2015).

Considered one of the greatest threats to public health, climate change is the term coined for the likely irreversible forces that are affecting everything from oceans and ice caps to hurricanes, droughts, and heat waves (McMichael et al. 2003, Solomon et al. 2009). While there are natural processes that have led to climate change events in our planet's history, the single largest contributor, especially in the past 200 years, is humans (Griggs and Noguer 2002, Crutzen 2006). A key indicator for quantifying anthropogenic changes over time has been the measurement of atmospheric greenhouse gases, most notably carbon concentrations in the primary forms of carbon dioxide (CO_2) and methane (IPCC 2013, US EPA 2016). Since the dawn of the industrial revolution (circa 1750), accumulated emissions of CO_2 concentrations in our atmosphere has increased from 280

parts per million (ppm) to over 400 ppm, a level our planet has not experienced in more than 3 million years (Lindsey 2020). Although CO_2 does not absorb as much heat per molecule as methane or nitrous oxide, it is more abundant and stays in the atmosphere longer (Lindsey 2020). Over time, the accumulation of mostly CO_2 and other greenhouse gases forms a "blanket" around the earth. As sunlight warms earth's land and ocean surfaces, thermal infrared energy is radiated. However, the growing "blanket" in our atmosphere essentially absorbs more of the radiated thermal energy than is reflected back out of the atmosphere, resulting in slight, but consistent increases in atmospheric temperatures. This phenomena is called the greenhouse effect (Mason 1989, Raval and Ramanathan 1989). Over the past century, global mean temperatures have increased 0.6°C and record breaking monthly and annual mean air temperatures are commonplace (Lindsey 2020).

Critical changes are also occurring in global precipitation and humidity, although less clearly. Some locations are experiencing increases in cumulative precipitation while others are decreasing (IPCC 2018). Over time, these changes will shift biomes, changing ecosystems, forcing species to either rapidly adapt or go extinct (Warren et al. 2018).

In addition to anthropogenic forces affecting changes in climate, human expansion, through deforestation and encroachment into new environments, is exacerbating negative effects to our planet's health (Myers et al. 2013, Bennett 2018). Interestingly, not all species are negatively affected from human effects on the planet. Infectious diseases, many of which our planet has had notable outbreaks of in past centuries, are poised to thrive (Patz 1996). In particular, diseases that are transmitted from animals to humans, zoonoses, are consistently experiencing new, ample opportunities to spill over into

human beings (Greer et al. 2008, Smith et al. 2014). In the past 80 years, the majority of global emerging infectious disease have been zoonotic (Jones et al. 2008, Smith et al. 2014, CDC 2018a)

While representing 17% of all infectious diseases, vector-borne diseases (VBDs) are increasing at a faster rate than any other zoonosis (Jones et al. 2008) and are a primary concern to public health (WHO 2017). Causing more than 700,000 annual deaths globally, vector-borne diseases are transmitted by many arthropod species, including mosquitoes, ticks, sand flies, black flies, Tsetse flies, midges, chiggers, mites, fleas, lice, and kissing bugs, among others (Gubler 1998, WHO 2017).

Among the long list of arthropod vectors, mosquitoes are among the most successful, in terms of adaptation and establishment of populations in new environments (Zheng et al. 2019). *Aedes albopictus* and *Ae. aegypti*, the primary vectors of Dengue, Japanese Encephalitis, Chikungunya, Yellow Fever, and Zika viruses, are considered among the top invasive species in the world (Bonizzoni et al. 2013), and are the archetypal species used in global examples of expanding vector-borne disease systems (Juliano and Philip Lounibos 2005).

Like all arthropods, the life cycle of mosquitoes are dependent on external abiotic forces (Madder et al. 1983, Knies and Kingsolver 2010). Increasing trends in temperature have accelerated mosquito development time, shortening the period from egg hatch to adult eclosion (Rueda et al. 1990, Alto and Juliano 2001). Rapid development times increase opportunities for a species to feed, mate, and lay eggs (Su and Mulla 2001, Foster and Walker 2002). Additionally, slight increases in annual temperature facilitate range expansions and establishments of native mosquitoes into new territories (Robinet

and Roques 2010, Ryan et al. 2018). Traditionally, the range expansion of notable key invasive mosquito species have been limited by overwintering temperatures (Jepsen et al. 2008, Bale and Hayward 2010). However, historic annual distribution maps of *Ae. albopictus* in North America show a slow, but consistent northerly expansion in the species distribution since its arrival in mid 1980s (Moore and Mitchell 1997, Armstrong et al. 2017, Ryan et al. 2018).

Prior to 1999, the primary vector-borne disease concerns in North America were Lyme disease and St. Louis encephalitis. In the summer of 1999, New York City experienced a sudden and large outbreak of West Nile virus (WNV) among birds and humans. In a matter of just 4 years, the virus spread from New York to California, occurring in all but 6 states. Largely attributing to its success, WNV is able to infect over 300 bird species and 60 mosquito species in North America (Hayes et al. 2005, Kramer et al. 2008, CDC 2017), providing ample opportunities to maintain presence in the environment as well as rapidly move vast distances.

West Nile virus is a single-stranded RNA virus from the family *Flaviviridae* (genus *Flavivirus*) (Gray and Webb 2014). The *Flaviviridae* family comprises the Japanese encephalitis virus serocomplex, and includes several viruses that cause encephalitis in humans: Japanese encephalitis, St. Louis encephalitis, and Murray Valley encephalitis viruses (Solomon 2004, Blitvich 2008). The virus was first reported in the West Nile district in the Northern Province of Uganda in 1937 in a 37-year old febrile woman (Hayes 2006). Phylogenic evidence supports that the strain that arrived in New York in 1999 (NY99) likely came from Israel, specifically the Isr98 strain (Lanciotti et al. 1999). However, it is not clear how the virus made its way to New York in 1999, but three

theories are most supported: 1. Infected mosquitoes hitchhiking in wheel wells of aircraft, 2. An infected avian pet was transported, or 3. Migrating avian species carrying the virus made their way to the Americas (Lanciotti et al. 1999, Murray et al. 2010, Hadfield et al. 2019).

Twenty years later, WNV is the etiologic agent responsible for the most mosquito-borne illnesses in humans in North America (CDC 2020a). Now endemic in the United States, WNV infections occur every year throughout all 48 conterminous states with outbreaks occurring in 2002, 2003, 2012, and 2018. In general, human cases tend to be highest in states with the most population: California, Colorado, Illinois, and Texas. However, incidence (per 100,000) are highest in the Dakotas and Nebraska (CDC 2018b).

The ability for WNV to infect subsequent mosquito vectors is largely dependent upon the capacity of avian hosts to amplify virus and mosquitoes species to be competent vectors. In North America, the most competent birds are from the family *Corvidae*, including Crows, Jays, Ravens, and Magpies (Komar et al. 2003). These species are highly susceptible and often succumb to infection. The virus depends on the ability for mosquitoes to acquire a blood meal in the short period during infection, but before death, when viremia is highest. However, studies have shown that a less susceptible and common avian host, the American robin (*Turdidae*), may provide adequate viremia for mosquito infection without succumbing to infection (Kilpatrick, Daszak, et al. 2006, Molaei et al. 2006, Hamer et al. 2008). Once a mosquito has acquired WNV through an avian blood meal, the survival of the virus is now dependent upon the vectorial capacity of the mosquito vector to infect subsequent suitable hosts (Kramer and Ciota 2015). The vectorial capacity (VC) is defined by the following equation: $VC = \frac{ma^2p^n}{-\ln(p)}$, where n is the

parasite's extrinsic incubation period (EIP, *n* days), *m* is the ratio of mosquitoes to humans, *p* is the mosquito's daily survival, and *a* is the human biting rate. Values further away from 1 demonstrate an increasing potential to transmit virus. Vectorial capacity varies by species and is influenced by environmental factors. Commonly reported mosquitoes with the highest vectorial capacities include *Culex (Cx.) pipiens, Cx. restuans, Cx. salinarius, Cx. tarsalis, and Cx. quinquefasciatus* (Turell et al. 2005). When an infected mosquito takes a blood a meal from a human or equine host, zoonotic spillover can occur. Coined "dead-end" hosts, infected humans and horses do not contain high enough levels of viremia to infect additional feeding mosquitoes (Lanciotti et al. 1999, Blitvich 2008). However, 75-80% of infected humans will be asymptomatic. The remaining 20-25% of humans develop mild to moderate febrile illness, called West Nile fever. About 1 in 150 infected humans, or less than 1%, will develop serious neuroinvasive illness that can result in death (CDC 2020a). Equine hosts are more susceptible to illness than humans, but a low-cost, widely available, and effective vaccine has been available since 2003, greatly reducing cases in the species (Ng et al. 2003).

In the Midwest U.S., *Culex pipiens* and *Cx. restuans* are the predominant WNV vectors and have adapted to living among urban and suburban environments (Hayes et al. 2005, Hamer et al. 2008). These environments have an abundance of trees, vegetation, and ornamental shrubs that provide nectar and resting sites for adult mosquitoes (Irwin et al. 2008). Additionally, these environments contain numerous artificial containers that have the potential to provide habitats for juvenile *Culex* mosquitoes (Vezzani 2007). For example, catch basins, water retention ponds and reservoirs, puddles (on impervious surfaces), blocked gutters, abandoned pools, bird feeders, and pots for outdoor plants,

have served as highly suitable habitats for *Culex* breeding (Geery and Holub 1989, Byrne and Nichols 1999, Harbison et al. 2014). In addition to these features, human behavior can facilitate and increase breeding environments. For example, discarded garbage, grass clippings, and leaves are common materials that can quickly impede water drainage in and around catch basins, leading to a covered and protected breeding habitat while simultaneously providing rich, organic substrate readily available for developing larvae (Mccall and Eaton 2001, Yee and Juliano 2006).

Efforts to control mosquito populations and monitor infection have been in place since the arrival of WNV. Mosquito control varies in the form of privatized businesses, city-based programs, county public health, and dedicated mosquito abatement districts (Nasci and Mutebi 2019). Largely dependent upon funding, some programs monitor and proactively treat for larvae and adults several months out of the year, while others are limited to a handful of personnel treating specific breeding sites after citizen complaints or evidence that infection is present (e.g., dead bird nearby, positive mosquito pool). Every state provides a "dead bird" hotline that allows for free submission and testing of specimens (CDC 2013), but public awareness and use has decreased since the initial outbreaks in 2002 and 2003. The effectiveness and ability in mosquito control's role to mitigate WNV in the environment has been debated, as numerous control agencies continuously report positive mosquito pools in treated locations, and human cases are still prevalent. However, millions of mosquitoes, many that are infected, are trapped and removed from the environment.

1.2. RATIONALE AND SCOPE

The epidemiology of vector-borne diseases are intertwined within the biologic interactions of the pathogen, arthropod vector, reservoir host, and in the case of WNV, the dead-end incidental human and equine hosts (Figure 1.1). These interactions result in constant changes and fluctuations in the life cycle of WNV from abiotic and biotic forces. Despite millions of dollars in resources and countless hours of person-time, mosquito control has not contained the virus in nature, but has had some success in reducing abundance of potential mosquito vectors. Cook and DuPage Counties (Illinois), compromising the greater Chicago area, are among the most well-funded and equipped locations to combat WNV. With four dedicated abatement districts and one of the largest public health departments in the county focused on mitigating mosquito populations in the city, the region has some of the most data-rich historic datasets on mosquito abundance and infection.

In Chicago, IL, the largest problem in attempting to control WNV has been the lack of understanding spatiotemporal patterns and ultimately, mitigating disease in humans. Specifically, indicators of WNV presence, through the forms of positive mosquito pools or dead birds, are not strongly associated with the locations of human cases. Efforts to address this problem have been attempted in my laboratory by previous students. In general, although strong overall, past predictive models poorly capture the magnitude of human infection in the locations where WNV is highest. In essence, because WNV is relatively rare in humans, there are many more locations that have never had WNV presence, creating an inflation of zeros. Most models evaluating WNV will be "strong" in their performances inherently as a result of this artifact – the models are very good at

predicting where cases will not be. The aspect of our work that will be most beneficial to public health – the successful prediction of both the location and magnitude of human cases – is immensely difficult to achieve.

In a generalized sense, surveillance and control programs base their operations on the ability to successfully remove potential mosquito vectors, and thus infection, from the environment. However, WNV has proven time and again to be resilient and avoid detection, in a surveillance and mosquito control sense, and continue to cycle in the environment. To begin to assist mosquito control and improve surveillance efforts in targeting of potential vectors, researchers should take a step back and reassess the tools at their disposal. Research that aims to improve our understanding of WNV disease dynamics should tailor efforts to reducing variance in this system. For example, the routine treatment of catch basins is a good approach, but should not be the sole location for targeting *Cx. pipiens* breeding sites. Additional strategies should focus on other breeding habitats that are being missed by treatment efforts. One possible method that can provide clues is to improve our evaluation in vectorial capacity, the index that quantifies the overall competency of a mosquito vector. Recalling the equation for VC, we find that no variables are static and can fluctuate based on location and time. For example, the average vectorial capacity for *Cx. pipiens* nationwide from 2001-2018 will be far less than the average vectorial capacity of the same species in Chicago, IL in the summer of 2012. Focusing research efforts that improve measurements of EIP, human biting rates, and the overall abundance in problematic areas at much finer temporal scales is essential for deciphering the nuances missed with traditionally applied surveillance

efforts. Although resource-consuming, these efforts may prove very beneficial in our understanding of how WNV maintains in any targeted environment at a given time.

1.3. OBJECTIVES

Since taking over the WNV modeling aspects in our lab, my primary focus was to improve upon earlier models, with the assistance of new approaches, analytical techniques, and supplemental data across ecological scales. The complex biological interactions of the WNV system – pathogen, mosquito, avian and human hosts – are mediated by a multitude of environmental and climatic forces. As climate change continues to influence VBD systems, our knowledge of these relationships must increase to better understand and ultimately, prevent human and animal illness.

The overarching goal of this dissertation is to create effective WNV disease models that can reliably forecast human infection in a given week, month, or year. Improvements to models, evaluated via statistical qualities, will inherently equate to a better understanding of the process of transmission. To achieve this goal, I have attempted to increase the spatial resolution of previously collected data as well as acquire new environmental, biological, and human-behavior based information. These efforts are stratified by previous human-risk levels, ranging from low to high, and have provided the finest-scale data specific to WNV transmission in the Midwest. Additionally, as part of the Midwest Center of Excellence in Vector-borne Diseases, all partners have provided a fortunate opportunity for increased collaboration and data-sharing. Taking advantage of these opportunities, this dissertation also evaluated the effects of WNV transmission across the entire 8-state partner region, incorporating all available mosquito infection and abundance data from 118 counties, ranging from 2000-2018.

The specific objectives of this dissertation are:

i. Determine how relationships of WNV may change under highly focused study regions in Chicago, IL. Previous modeling efforts in Chicago have evaluated WNV at the county-level. We hypothesize that by reducing the extent of our study area and collecting additional fine-scale data, we will capture more variance in WNV transmission and increase overall model strength. The main questions investigated for this objective are: 1. How do the relationships of key variables in smaller scales compare to that of the larger Cook/DuPage model? 2. What are the similarities and differences among overlapping locations that were predicted accurately by the larger model and those that were predicted poorly? 3. Did the predictive power in these locations improve with fine-scale models? 4. What effect did newly acquired data have on prediction of human WNV illness?

ii. Evaluate the strengths and weaknesses in WNV model performance across scales and determine what conditions are optimal for future investigators, control personnel, and public health officials. I hypothesize that as spatial scale increases, model performance decreases, the number of covariates increases, and the magnitude in effect of each covariate will decrease. The main questions investigated for this objective are: 1. How does the overall importance and relationships of covariates change for each model by scale? Are there ecological factors that contribute to changes in covariates by scale? 2. How do best-fit models at one scale perform when applied to another scale? 3. Overall, what scale is optimal for evaluating WNV infection?

iii. Increase understanding in the spatiotemporal relationships of WNV throughout the midwestern United States. We hypothesize that a standardized analysis of all available mosquito records will provide good predictions of WNV infection where resources are highest and poor predictions where surveillance efforts are lowest. The main questions investigated for this objective are: 1. Where are mosquito surveillance efforts the most prevalent and where are they most lacking? Does this correlate to increases (or decreases) in mosquito infection? 2. What are the main drivers of WNV infection in mosquitoes in the Midwest? 3. How do relationships of covariates change by county? 4. Can we generalize mosquito infection across the Midwest and predict when and where future hotspots will occur?

1.4. FIGURE

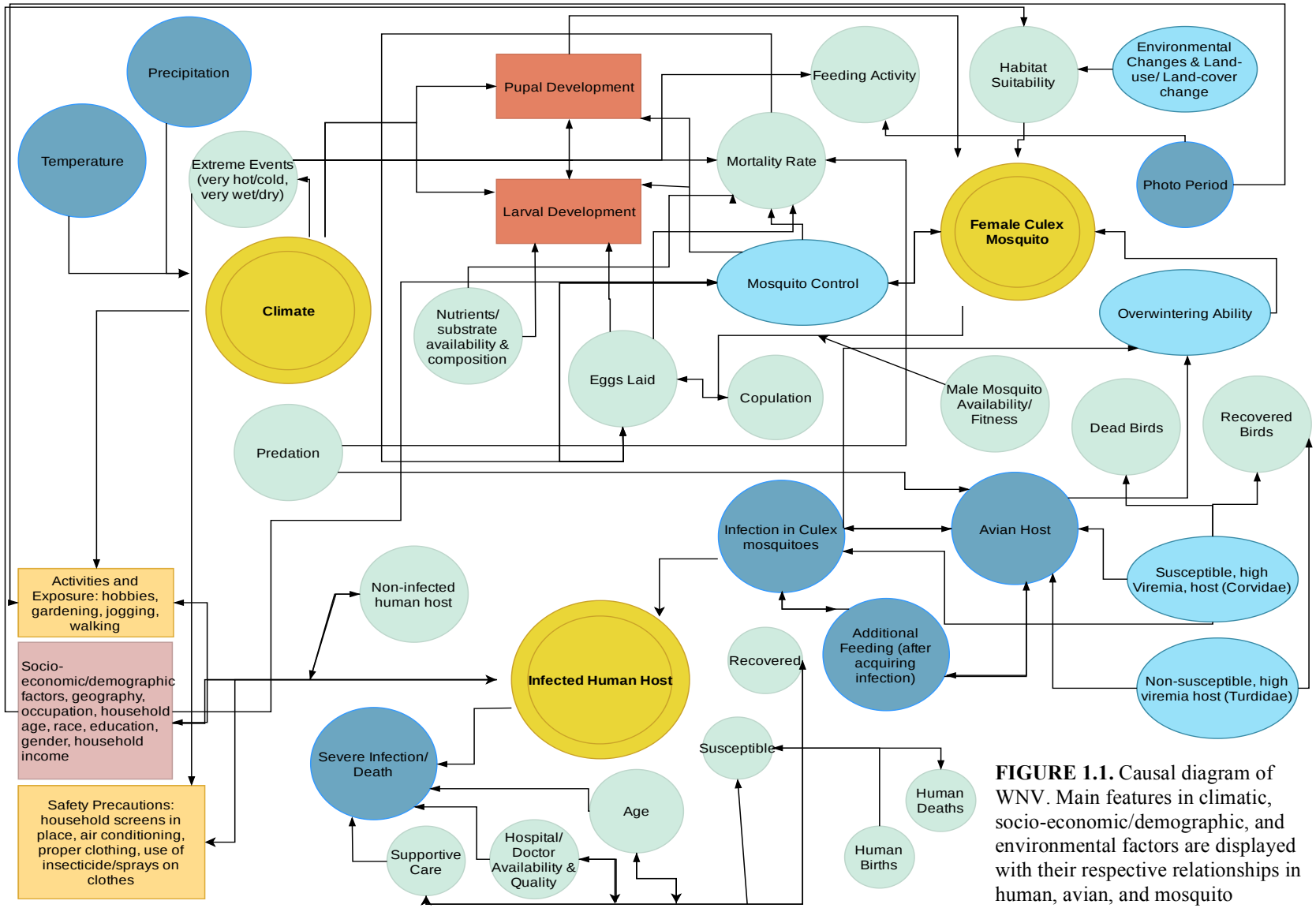


FIGURE 1.1. Causal diagram of WNV. Main features in climatic, socio-economic/demographic, and environmental factors are displayed with their respective relationships in human, avian, and mosquito biology.

CHAPTER 2: ASSESSING ULTRA-FINE-SCALE FACTORS TO IMPROVE HUMAN WEST NILE VIRUS DISEASE MODELS IN THE CHICAGO AREA

2.1. ABSTRACT

Background: Since 1999, West Nile virus (WNV) has moved rapidly across the United States, resulting in tens of thousands of human cases. Both the number of human cases and the level of mosquito infection (MIR) vary across time and space and are related to numerous abiotic and biotic forces, ranging from differences in microclimates to socio-demographic factors. Because the interactions among these multiple factors affect the locally variable risk of WNV illness, it has been especially difficult to model human disease risk across varying spatial and temporal scales. Cook and DuPage Counties, comprising the city of Chicago and surrounding suburbs, are among the most affected by WNV in the United States. Despite active mosquito control efforts, there is consistent annual WNV presence, resulting in more than 285 confirmed WNV human cases and 20 deaths in the past 5 years in Cook County alone.

Methods: A previous WNV model for the greater Chicago area identified the fifty-five most high and low risk study areas in the Northwest Mosquito Abatement District (NWMAD), an enclave ¼ the size of the previous study area. In these locations, human WNV risk was stratified by strength of predictive success, as indicated by differences in studentized residuals. Within these areas, an additional two-years of field collections and data processing was added to a 10-year WNV dataset and assessed by an ultra-fine-scale multivariate logistic regression model.

Results: Multivariate statistical approaches revealed that this ultra-fine-scale model resulted in fewer explanatory variables while improving upon the fit of the existing

model. Beyond mosquito infection rates and climatic factors, efforts to acquire additional covariates only slightly improve model predictive performance.

Conclusions: These results suggest human WNV illness in the Chicago area may be associated with fewer, but increasingly critical, key variables at finer scales. Given limited resources, this study suggests a large variation in the significance to model performance, and provides guidance in covariate selection for optimal WNV human illness modeling.

2.2. INTRODUCTION

In December of 1937 in Northern Uganda, a 37-year-old woman became ill with a fever of 100.6°F (Smithburn et al. 1940). She would later become the first documented human infected with the West Nile virus (WNV; Family *Flaviviridae*), a mosquito-borne disease first reported in the West Nile region of Uganda. West Nile virus first arrived to the United States (U.S., New York, NY) in 1999, most likely via a hitchhiking infected mosquito in an airline wheel well (Hadfield et al. 2019). This newly introduced WNV strain matched the Isr98 strain, isolated from a single goose in Israel in 1998 (McLean et al. 2002). Once arriving in New York, the virus took only three years to traverse the contiguous U.S., reaching California in 2002 (Sejvar 2003). The virus has now become one of the most widespread arboviruses in the world, and is present in every continent except Antarctica (Kramer et al. 2008).

In the midwestern U.S., mosquitoes of the *Culex* (*Cx.*) genus are the main vectors for transmitting WNV (Goddard et al. 2002). *Culex* mosquitoes are capable of feeding on several hosts to satisfy one blood meal, increasing the opportunity for multiple infections across species (Hamer et al. 2009). Although primarily ornithophilic, prior studies

indicate that *Cx.* species may shift feeding preferences to humans later in the summer months (Kilpatrick, Kramer, et al. 2006, Russell and Hunter 2012). Humans and other mammals, most notably horses, are considered "dead-end" hosts, not capable of producing sufficient levels of viremia to subsequently infect biting mosquitoes (Bowen and Nemeth 2007).

From 1999-2018, there have been a reported total of 50,830 human cases resulting in 2,330 deaths across the US (Centers for Disease Control (CDC) 2019). In many cities and states that experience high WNV incidence, there are efforts in place to control mosquito populations. However, despite these methods, WNV continues its epizootic and enzootic cycles year to year, and large-scale outbreaks have occurred in the years 2002, 2003, and 2012 across the United States. At local scales, drivers of human disease, including WNV, vary in actual effect and magnitude from that observed on state, regional, or national scales. Previous studies have identified common abiotic and biotic factors associated with human WNV illness, including prior weather conditions (weekly temperature and precipitation lags), mosquito infection and abundance, socio-demographic characteristics of the local population, and level of public awareness and education, but these were all at state or regional scales (Ruiz et al. 2004; Kilpatrick and Pape 2013; Manore et al. 2014; Roiz et al. 2014; Rosà et al. 2014; Wimberly et al. 2014; Hahn et al. 2015; Giordano et al. 2017).

Karki et al. (2020) is one of the few studies to evaluate weekly spatiotemporal factors and their associations with human WNV illness at a smaller scale, in a highly urban 2-county area (Cook/DuPage counties) that includes the greater Chicago, IL area. This region consistently experiences one of the highest annual WNV incidences in the country.

The previous study incorporated among the finest temporal and spatial scales known to date, using 1-km hexagon grids to minimize biases from political boundaries. While an excellent overall model fit was achieved by using a large number of explanatory variables, the relative importance of covariates and the resulting disease prediction across mico-scales is still not understood.

The Northwest Mosquito Abatement District (NWMAD), occupying the northwest corner of Cook County, is one of Chicago's four abatement districts responsible for mosquito control. The NWMAD also has excellent long-term mosquito abundance and testing data throughout its jurisdiction, allowing for evaluations and comparisons of mosquito WNV prevelance with regard to minimum infection rate (MIR) and vector index (VI). This study targeted fifty-five individual 1-km hexagons within the NWMAD for an ultra-fine-scale (UFS) assessment of human WNV illness spatio-temporal variability in suburban environments. Specifically, this study's main objectives were to: (i) evaluate and contrast key variables in this study to the larger Cook/DuPage model, (ii) assess the similarities and differences among locations that were predicted accurately by the larger model and those that were predicted poorly, (iii) quantify the impact of newly acquired data on prediction of human WNV illness, and (iv) determine if vector index is a stronger predictor than the additive effects of MIR and mosquito abundance. Ultimately, this study aims to highlight how WNV disease variance may be better captured at finer spatio-temporal scales. The results of this study will provide future researchers, public health agencies, and abatement districts essential details and suggestions for improving WNV prediction and optimizing efficiency of targeted mosquito control efforts.

2.3. METHODS

This project was approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign, the Illinois Department of Public Health (IDPH), and the University of Illinois Biosafety Committee. Human case data were provided by IDPH without any personal identifying information.

2.3.1. Study area

This study was conducted within the NWMAD, a 605-km^2 area that comprises the northwest suburbs of Chicago (Cook County, IL, Figure 2.1). As described in Karki et al. (2020), all model data were summarized and processed within 1-km diameter hexagons, as a neutral configuration in both size and shape, free of any political boundaries. Using statistical selection processes (described below), fifty-five of the 1,019 hexagons within the NWMAD were selected as the observational units for this study.

2.3.2. Model covariates

The Cook/DuPage model evaluated forty covariates derived from a variety of abiotic and biotic factors, including climate and weather records, mosquito infection, socio-demographic census data, and other biological conditions (described below). For this study, additional data processing and field collections resulted in forty-two additional non-collinear independent variables (Table 2.1). Each variable was independently calculated by hexagon for CDC epidemiological weeks 18-38 (Sunday-Saturday) of the years 2005 through 2016 (CDC 2019).

2.3.2.1. Previously existing data

2.3.2.1.1. Human illness

Human WNV cases in Illinois were classified as either confirmed[1] or probable[2], as reported to the IDPH by public health or licensed medical professionals (reporting of WNV cases is required in the state). I recognize that exposure to mosquito-borne disease occurs often and in many locations. Confirming the moment an infected mosquito inoculates a human is nearly impossible to document. Therefore, I assumed human cases were exposed to WNV at their home addresses. The latitude and longitude point locations were provided to the third decimal degree and aggregated to the hexagon level for analytical and display purposes. Human cases were converted into binary form (presence/absence of illness) and weekly case rate, controlling for human population, for each hexagon.

2.3.2.1.2. Abiotic Predictors

Land Cover: The 2011 United States Geological Survey (USGS 2011) National Land Cover Database (NLCD) provided 30 m resolution classified raster data for the NWMAD. The raster comprising NWMAD was clipped, extracted, and tabulated by landscape code using the tabulate area tool in ArcGIS 10.5.1 (Environmental Systems Research Insititute 2011). There were 15 unique land cover types: forests (deciduous, evergreen, and mixed), urban (developed open space, developed low intensity, developed medium intensity, and

[1] The case definition for a confirmed case of arboviral encephalitis in Illinois is a clinically compatible illness that is laboratory confirmed at a public health laboratory. The laboratory criteria are a fourfold or

[2] A probable case of arboviral encephalitis is a clinically compatible illness occurring during the season when arbovirus transmission is likely to occur and with the following supportive serology: a stable (twofold or smaller change) elevated antibody titer to an arbovirus, e.g., > 320 by hemagglutination inhibition, > 128 by complement fixation (CF), > 256 by IF, > 160 by neutralization, or a positive serologic result by enzyme immunoassay (EIA) or MAC ELISA.

developed high intensity), open water, herbaceous wetlands, cultivated crops, wetlands (woody and herbaceous), grassland, barren land, and shrubs. Proportions of each type within each hexagon were calculated using the 30 m raster resolution.

Weather: Daily mean temperature and precipitation were acquired from the PRISM Climate Group (Oregon State University 2019), provided as 4-km resolution grids. Weekly mean temperatures were calculated by taking the average of each of the seven days of the week, whereas weekly precipitation totals were calculated as a sum of each of the seven days of the week. As a proxy for winter temperature, the monthly average for each January from 2005-2016 was also calculated. Using the zonal statistics as table function in ArcGIS, each mean temperature and precipitation value was extracted for each hexagon in this study.

2.3.2.1.3. Biotic Predictors

Mosquito infection: All mosquito infection data were acquired from the Illinois Department of Public Health (IDPH), the state agency responsible for collecting and maintaining standardized mosquito collection and testing data. Mosquito infection is defined as the minimum infection rate, calculated by the following equation:

$$\frac{\#\ of\ positive\ mosquito\ pools}{total\ specimens\ tested} \text{ x } 1000,$$

where a mosquito pool in this analysis consisted of up to 50 female *Culex* mosquitoes that were collected by the same trap. A vast majority of the tests used to identify the presence of WNV was the Rapid Analyte Measurement Platform (RAMP), although some mosquito pools were also tested by Real Time reverse transcriptase polymerase chain reaction (RT-PCR) or VecTest.

Trap locations were provided by the IDPH. Whenever precise spatial locations were not available, the existing address on file was used to generate a geocoded trap location. The MIR values for each trap were calculated and interpolated across the NWMAD by inverse distance weighting (IDW) in ArcGIS. The average MIR values were extracted for each hexagon by using the zonal statistics as table function in ArcGIS.

Demographic: Total population and racial composition (White, African American, Hispanic, and Asian) at the census block level were extracted from the 2010 U.S. Census then converted as a percentage for each hexagon. Additionally, age of housing (built before 1940, 1940-1969, 1970-1989, and post 1990) and income were averaged for each hexagon using data provided by 2015 American Community Survey. These data were processed in ArcGIS using the intersection tool.

2.3.2.2. Newly added data

2.3.2.2.1. Abiotic Predictors

Catch basin density: The NWMAD provided point data for each catch basin within its jurisdiction. All point data were then aggregated to each hexagon using the spatial location join feature in ArcGIS. A combined total of 8,443 catch basins were recorded among all hexagons (min = 1, max = 543).

Size and distribution of commercial and residential lots and buildings: High-resolution (1 m) aerial imagery from ArcGIS and USDA (2018) were used as a basemap for each hexagon. Each permanent structure (e.g. residence, shed, garage, deck) was traced and converted to polygons in ArcGIS. The area and perimeter of each polygon was calculated and aggregated for each hexagon. Commercial and residential lots were provided by

Cook County Data Catalogy (2019), using 2016 tax appropriations. In total, there were a combined 22,892 lots with 24,468 buildings or permanent structures.

Light pollution: Radiation from light pollution was provided by the New World Atlas of Artificial Night Sky Brightness (Falchi, F., Cinzano, P., Duriscoe, D., Kyba, C. C. M., Elvidge, C. D., Baugh, K., Portnov, B., Rybnikova, N. A., Furgoni 2016a, 2016b). Estimates of light pollution were acquired from 2014 data of the VIIRS DNB sensor on the Suomi National Polar-orbiting Partnership (NPP) satellite. Pixel resolution was 0.75 km; mean value for each 1-km hexagon was calculated in ArcGIS.

2.3.2.2.2. Biotic Predictors

Historical mosquito abundance: The NWMAD consistently collected and diligently maintained their mosquito trapping and identification data throughout the study period. Once deployed, traps were usually checked at least twice per week. Over the 2005-2016 study period, there were a total of 59 traps used in the NWMAD, resulting in a total of 48,406 female *Cx.* spp. from 22 light traps, and 1,110,024 from 37 gravid traps. Weekly mosquito collections by trap were geocoded and interpolated across all hexagons via IDW and extracted using the zonal statistics as table function for each hexagon in ArcGIS 10.5.1. Mosquito abundance was calculated as the weekly cumulative number of captured female *Culex* spp. from each respective gravid trap (GT) and light trap (LT). Since early trap data did not reliably identify mosquitoes to species, all *Cx.* spp. values were pooled. However, prior studies from the Chicago region indicated that *Cx. pipiens* and *Cx. restuans* are the major *Cx*. species present in this area. Normalized Difference Vegetation Index (NDVI): To evaluate the magnitude of all vegetation, NDVI was incorporated by hexagon, recorded as an average value at three time points of each year: CDC

epidemiologic weeks 21 (3rd-4th week of May), 28 (2nd-3rd week of July), and 35 (4th week of August-1st week of September). These CDC epidemiologic weeks mark the center of each the three 8-week active WNV periods in the Midwest, represented as T1 = low WNV activity, T2 = high WNV activity, and T3 = moderate WNV activity. The best available Landsat 7 or 8 bands for each respective time period were acquired from EarthExplorer (USGS 2019) and processed in ArcGIS 10.5.1.

Human exposure during crepuscular time periods: Human activity observations were conducted in public spaces inside each hexagon, during the crepuscular hours between 6-9:30pm, the preferred feeding period for *Cx. pipiens/restuans* (Caglar et al. 2003). Observations were conducted within each hexagon for a total of ten minutes per visit. Specifically, a researcher remained stationary for 2 minutes, walked 2 minutes, remained stationary in the new position for 2 minutes, walked back to origination point for 2 minutes, then remained stationary in the original position for 2 final minutes. Human exposure was determined as any period in time a person was outside of any building, vehicle, or enclosed dwelling during the observation period. Observations were classified by apparent gender and age category (child, adult, or senior citizen).

Human Landing Catches: During human observations, another researcher collected human-seeking mosquitoes via the human landing catch (HLC) method at the same location. Each HLC visit exposed the researcher for fifteen minutes. To mitigate actual biting events, the researcher would expose only one limb (arm or leg) at a given time. Any mosquitoes that landed would be collected via mechanical aspirator and transferred to a collection vial. All collected mosquitoes were transported to the NWMAD within 2 hours and stored at -80°C. All mosquito specimens were identified to species within three

days. Any mosquitoes identified as *Culex* spp. were sent to the Fritz Lab at the University of Maryland for species confirmation by *Cx. pipiens* group-specific primers via PCR.

Vector Index: The vector index was calculated as an estimate of the relative number of WNV-infected mosquitoes. Specifically for this study, VI was calculated as the average number of pooled *Culex* spp. collected per trap-week multiplied by the proportion of mosquitoes infected with WNV. The following equation was modified from the CDC (2013):

$$\text{VI} = \sum_{i=Culex\ \text{spp.(pooled)}} \overline{N}_i \widehat{P}_i,$$

where $\overline{N}_i$ = average density (number of mosquitoes per trap week) and $\hat{P}_i$= estimated MIR (proportion of mosquito pools testing positive for WNV). Calculated weekly VI for each trap by week was then interpolated via IDW method for estimations across the NWMAD.

Nuisance Factor and Human WNV Added Risk: Mosquitoes collected via HLC methods were categorized into one of two types: nuisance and WNV vectors. Since the majority of mosquitoes collected were non-*Culex*, a quantitative index, nuisance factor, was created to provide a risk spectrum of encountering nuisance mosquitoes in a given hexagon. The following equation defines the nuisance factor:

$$Nuisance\ Factor = \frac{\frac{Human\ Observations}{Hour} * \frac{Nuisance\ Mosquitoes}{Hour}}{100}$$

Nuisance factor values ranged from a low of 0 to a high of 32.3. To quantitatively estimate potential risk for exposure to disease within a given hexagon, the human WNV added risk factor was also created. This index is defined by the following equation:

$$Human\ WNV\ Added\ Risk = \frac{\frac{Human\ Observations}{Hour} * \frac{Culex\ \text{spp.}}{\text{Hour}}}{100}$$

Human WNV added risk ranged from a low of 0 to a high of 1.44.

2.3.3. Statistical methods

2.3.3.1. Location selection

Of the total 1019 hexagons within the NWMAD, we selected fifty-five (5.4%) as the maximum number of sites that our research team could visit for fifteen minutes each, weekly. The subset of fifty-five hexagons were selected based on two criteria: (1) the size of the human population was > 0, and (2) where the previous Cook/DuPage model either predicted human WNV extremely well or extremely poorly, as determined by the 2005-2016 average residual output. Furthermore, the residual output was stratified by those locations that had or had not experienced a human case during the 12-year period. These processes created a performance spectrum consisting of five categories of hexagons: negative residuals without a human case (NR0), low residuals without a case (LR0), low residuals with a case (LR1), positive residuals without a case (PR0), and positive residuals with a case (PR1) (Table 2.2). No hexagons with negative residuals in the Cook/DuPage model had experienced a human case. The spatial arrangements of these hexagons provide adequate coverage of the NWMAD's jurisdiction (Figure 2.2).

2.3.3.2. Model Selection

Two seasons of field collections and processing of new data provided the UFS model with an additional 42 covariates not made available in the previous Cook/DuPage model. The generation of linear and logistic regression models began with a two-step selection process for the initial covariate inclusion: (1) conduct a univariate analysis with each predictor (independent variable) to the WNV disease outcome (binary = logistic, case rate = linear, dependent variable). Candidate variables for multivariate analysis were selected

using slightly more conservative p-value than Bursac et al. (2008), p-value ≤ 0.20 vs. ≤ 0.25). Models that create cut-off values of p-value ≤ 0.1 for purposeful univariate covariate selection can erroneously prevent important variables from entering final models (Bendel and Afifi 1977; Greenland and Mickey 1989); (2) the final model, a generalized linear model personality with a Poisson distribution and probit link function, was selected using forward selection method, selecting the final model based on the Bayesian information criterion (BIC). Non-significant covariates were removed from the final model as a product of the iterative selection process. Secondarily, a receiver operating characteristic (ROC) curve was used to visualize overall model performance and Area Under the Curve (AUC) was calculated. All predictors were evaluated for multicollinearity using the PROC REG procedure (SAS Institute Inc. Cary, NC, USA). Regression analyses were analyzed using the Fit Model feature in JMP 14.2.0 (SAS Institute Inc. Cary, NC, USA). Binary WNV case outcome was analyzed as a nominal logistic personality. The continuous WNV case rate outcome was analyzed as a standard least squares personality.

2.3.3.3. Model Comparisons

Human WNV illness in the NWMAD was assessed under four model environments, each expressing a defined set of specific parameters. The four model environments were:

1. MIR & Mosquito Abundance (contains no VI covariates),
2. Vector Index (contains no MIR or mosquito abundance covariates),
3. Best-Fit (best fit with all covariates in respective assessment), and
4. Global (all covariates made available in respective assessment)

As a comparison, the original Cook/DuPage model was fit using only the 40 covariates included in the final model fit from Karki et al. (2020). Each of these four model environments were assessed using four different covariate sets:

1. All covariates (82 available covariates),
2. Excluding HLC and human observations covariates (74 available covariates),
3. Force-fitting HLC and human observations covariates (8 forced covariates, 82 available covariates), and
4. Only the covariates made available to the Cook/DuPage 2019 model (control model, 40 available covariates).

Under each model environment and covariate set, the outcome of human WNV illness was analyzed using:

1. Logistic regression (presence/absence human WNV illness) and
2. Linear regression (WNV case rate) methods.

In total, there were 36 models assessed; models are named using the convention $E_xC_yO_z$, where x is the model environment number (0-4, with number 0 assigned to the control environment), C is the covariate set number (1-4), and O is the outcome number (1-2). For both logistic and linear regression, each of the four model environments was fit using each of the four covariate sets. In addition, the control models using only the covariates from the final Cook/DuPage model applied to the UFS region were fit with and without force fitting HLC and human observation covariates.

Half of the models were assessed under logistic and linear outcomes, respectively, and based on the *# of Significant Covariates* (quantity of variables included in final model with $p<0.05$) and *Degrees of Freedom* (the number of values in the final model that are

free to vary). Overall model performance was determined by BIC. While BIC and Aikake's Information Criterion (AIC) are both maximum likelihood estimators, BIC was chosen to determine model strength due to its stronger penalty term for covariate inclusion (Schwarz 1978).

2.3.3.4. Covariate Performance

Similarly to the model performance index, to evaluate the performance for all covariates across 18 logistic and 18 linear models, each of the 82 covariates were standardized by creating the following index:

$$\bar{p}_{\text{Covariate}} = \frac{Significance\ Level}{Data\ Availability}$$

where: *Significance Level* = significance level of covariate in each of the 36 final models ($p<0.001 = 4$, $p<0.01 = 3$, $p<0.05 = 2$, included in the final model = 1), and *Data Availability* = tradeoff between resources required to acquire a respective covariate (level 1 = data widely available, no processing needed, level 2 = data available, requires minimal to moderate processing/analyses, level 3 = data available, requires extensive processing/analyses, level 4 = data not available, needs to be collected, processed, and analyzed, Appendix A, Table 1). The final net prediction: availability tradeoff used to create the Data Availability variable are categorical and based on the author's personal experiences with data used in this study.

2.4. RESULTS

2.4.1. Location Description

The UFS study area contained a total of forty human WNV cases from 2005-2016 (Table 2.2).

2.4.2. Model Fitting

With the exception of model $E_4C_30_1$, all models successfully converged (Tables 2.3 & 2.4), with AUC for the logistic models ranging from 0.84 to 0.97 and BIC values of 576 to 769, while BIC values for linear regression models ranged from -227444 to -181982. Despite converging, all global models (n=8) were excluded from the analysis due to statistical overfitting.

2.4.3. Model Comparison

The highest performing WNV human risk models were $E_3C_4O_3$ (Cook/DuPage Best Fit, df = 8, BIC = -227444) and $E_2C_4O_1$ (Cook/DuPage + VI, df = 14, BIC = 576.2), for linear and logistic regressions, respectively (Tables 2.3 & 2.4).

The top five models that predicted human WNV cases strongest were represented by the control (E_0, n=2), best-fit (E_3, n=2) and vector index (E_2, n=1) environments (Figure 2.3B, Table 2.5). These models' corresponding covariate sets were represented by variables only available to the original Cook/DuPage models (C_4, n=4), and force-fitting HLC covariates (C_3, n=1) environments.

2.4.4. Covariate Performance

Of the 82 available covariates, 70 (85.4%) were included at least once among a given model, excluding the overfit global models. Of the 41 covariates (58.6%) that were greater than the mean covariate performance, seven were highly efficient (determined by natural break in the distribution), providing a crude estimation as most valuable variables for human WNV estimation (Figure 2.3A). These covariates are provided here in descending order of most importance: tempc (temperature (°C), $\bar{p}$ = 1.15), preci (precipitation (mm), $\bar{p}$ = 1.14), Yr (year, $\bar{p}$ = 1.0), templag3 (temperature lagged by 3

weeks, $\bar{p}$ = 0.92), blpct (barren land (%), $\bar{p}$ = 0.92), precilag1 (precipitation lagged by 1 week, $\bar{p}$ = 0.90), and VIlag4 (vector index 4 weeks prior, $\bar{p}$ = 0.88). All eight HLC and human observation covariates were included in a final model, but none performed highly ($\bar{p}_{\text{each HLC Covariate}}$ = 0.25). Estimates and calculations for individual covariates are available in Additional file 1.

The eight HLC and human observation covariates provided significant differences in observations and mosquito collections by hexagon type (Figures 2.4A & 2.4B). The indices, nuisance mosquito exposure and human WNV added risk, significantly differed by hexagon type (Figure 2.4C). Hexagons designated as PR1 (positive residual (underpredicted actual cases) with a prior human WNV case) were found to have the most human observations and collected mosquitoes (from both *Culex* and non-*Culex* spp.) per visit. This combination of factors provides hexagons among the PR1 type as the most "risky" in regard to human WNV added risk and increased nuisance mosquito exposure (Figure 2.5).

2.5. DISCUSSION

With the exception of $E_0C_4O_2$, the Cook/DuPage control models, in conjunction with all other covariate and outcome sets, were consistently ranked moderate to low in WNV predictability and net value. Despite excellent prediction capabilities for the larger Cook/DuPage counties study area, this finding suggests that the UFS study areas have more variance from unaccounted sources that are missed or oversimplified in traditional, large-scale models.

In addition to model comparisons, this study evaluated the performance of the newly acquired VI in comparison to the previously used MIR in combination with mosquito

abundance. The original Cook/DuPage model only used MIR and its associated 4-week lags and achieved very good prediction results over the 2-county area. Overall, when fit to the UFS study area, adding mosquito abundance and associated 4-week lags improved this model. When evaluating WNV prediction as a linear outcome, the best-fit model using only covariates available to the original Cook/DuPage model was the highest performing in WNV predictability. However, when evaluating WNV prediction as a binary outcome, VI (a product of MIR and abundance together) and its associated 4-week lags replaced MIR as the best predictor of human WNV. While no model emphasizing MIR and abundance was selected as one of the best predictive models, at least one of these variables (and their associated lags) were represented in 4 of the 5 best models (control and best-fit, n=2 for each model). On the contrary, VI, as an emphasized model environment, was selected as the best performing logistic model. Both MIR and VI are critical components in predicting WNV. Deciding between the two biological indicators will be largely dependent upon the data availability for each model of interest. However, if resources are limited, the net model value leans in favor of using MIR.

The addition of 42 new covariates required a significant allocation of resources but provided minimal benefits towards reducing variance in human WNV prediction. Fortunately, this study suggests that excellent disease prediction models can be achieved with conventional covariates that are publicly available, requiring little to no processing and/or analyses (data availability scores ≤ 2, Figure 2.3B). However, any covariate used should be adjusted and properly designed for the highest spatial and/or temporal resolution possible, which may require additional efforts to accomplish.

Extensive review of literature indicated no other studies have evaluated covariate strength given limited resources, particularly in the context of making decisions to acquire data. Therefore, the categorizations of covariates by resource allocation [values ranging from 1 (low) to 4 (high)] are based on the experiences of the author during this study. These values are subjective and may vary across institution or research group, but they may be used as a general estimation in model selection and decision-making. For example, variables related to building and lot size (avg. bldg. area: avg. lot area, bldg. footprint area avg., bldg. footprint area total, bldg. footprint peri avg., bldg. footprint peri total, and total bldg. area: total lot area) were all ranked a value of 4 because of extensive data processing and review. The author downloaded high resolution, cloud-free satellite images that were used as a basemap for digital tracing of every building structure (houses, businesses, sheds, detached garages, storage units, etc.) and lots (residential and commercial). This resulted in >47,000 structures and lots digitally traced manually. On the other hand, weather variables (e.g. preci, tempc) were ranked a value of 1 because very little resources were devoted to have the data in a "ready" state. The source of these data, PRISM Climate Group, allows for monthly summaries to be downloaded and extracted with one quick geostatistic process.

This study also aimed to address a key missing index that few studies have evaluated: the relationship of human activity, mosquito exposure, and WNV disease risk. While the related variables did not greatly impact overall model strength, they did provide key insight into a potential key in WNV ecology – the areas that were previously underpredicted with recorded human WNV (hex type: PR1) were consistently found to have the most human activity at crepuscular times, the most mosquitoes overall, and the

most *Culex* mosquitoes. However, our results appear to contradict the findings of Read et al. (1994), who discovered that as reports of biting nuisance mosquitoes increased beyond 2 per minute, outdoor human activity rapidly declined. Our results indicate that as mosquito collections increased, human observations also increased (Figure 2.4). Not only is this a potentially dangerous combination that can foster environments ideal to mosquito-human spillover, previous modeling efforts failed to capture these cases. Future directions will target these highly susceptible locations and aim to capture any additional unaccounted variance.

Like all disease modeling efforts, there are always reporting biases that directly affect true case prevalence. Unfortunately, many vector-borne diseases are largely underreported (Bowden et al. 2011, Nelson et al. 2015, Waterman et al. 2015, CDC 2018), as human cases are vastly overlooked or misdiagnosed, largely due to low severity in disease manifestation in a majority of cases (CDC 2015, Rosenberg et al. 2018). This creates difficulties in predicting when and where VBD incidence will arise. In the Chicago area, models in both the UFS and Cook/DuPage locations have very good human WNV prediction capabilities. Despite having among the highest total number of human WNV cases in the U.S. [CDC National arboviral surveillance system (ArboNET, CDC 2020)], this region has more observational units denoted as non-cases than cases. That has resulted in models with excellent accuracy in predicting where there are no human cases, thus inflating the true accuracy of our models. Nonetheless, while our models are able to reliably predict where human cases are present, the magnitude of effect can be missed (e.g. “hot spots” with greater than 1 case may not be represented).

Disease modelers need to be cognizant of saturating their efforts, both statistically and biologically. Statistically, additional and meaningful covariates will usually improve model fit parameters. However, the inclusion of too many variables can result in overfitting, resulting in models failing to converge (Babyak 2004, Hawkins 2004, Lever et al. 2016). It is estimated that about 80% of human WNV infections are unreported, as clinical signs are minor or asymptomatic (CDC 2010, Petersen, Brault, et al. 2013). The remaining 20% of humans develop West Nile fever, and among this group, about 1% will develop severe and sometimes fatal neuroinvasive disease. It is possible that no matter the amount of effort to improve model fit, there is an element of variability attributed with infected humans not seeking medical attention and thus, reducing disease prevalence (Petersen, Carson, et al. 2013).

Overall, when compared to the Cook/DuPage model, the best UFS models required fewer predictors and produced a stronger overall fit using most, if not all, the same covariates made available to both model types. Spending the resources (time, money, human-power, processing, analyses, logistic, etc.) to acquire additional covariates may not necessarily be worth the impact on improving human WNV modeling predictions. Rather, fine-tuning the traditional covariates (climatic, weather, and MIR, for example), to the highest spatiotemporal resolution possible may be the most efficient use of resources to minimize variance in VBD prediction models.

2.6. TABLES AND FIGURES

Table 2.1. List of covariates used previously in Cook/DuPage Counties WNV model and those newly acquired variables used in newly revised 55 hexagon UFS model.

Covariate Information					
Designation		**Description**	**Notation**	**Cook/DuPage Model**	**Ultra-fine-scale Model**
Environmental	Land Cover	Proportion of developed open space	dospct	X	X
		Proportion of developed low intensity	dlipct	X	X
		Proportion of developed medium intensity	dmipct	X	X
		Proportion of developed high intensity	dhipct	X	X
		Proportion of deciduous forests	dfpct	X	X
		Proportion of evergreen forests	efpct	X	X
		Proportion of mixed forests	mfpct	X	X
		Proportion of barren land	blpct	X	X
		Proportion of shrubs	shrubspct	X	X
		Proportion of grassland	glandpct	X	X
		Proportion of pasture	pasturepct	X	X
		Proportion of cultivated land	clpct	X	X
		Proportion of woody wetlands	wwpct	X	X
		Proportion of herbaceous wetlands	hwpct	X	X
		Proportion of total forest	ftotpct		X
		Proportion of total wetlands	wtotpct		X
		Proportion of open water	owpct	X	X
		Normalized Difference Vegetation Index	NDVI		X
Biological	Minimum Infection Rate (MIR)	MIR one week before	mirlag1	X	X
		MIR two weeks before	mirlag2	X	X
		MIR three weeks before	mirlag3	X	X
		MIR four weeks before	mirlag4	X	X
		Average MIR current week	MIRmean		X
		Difference in weekly average MIR from 12-year average	MIRdiff		X
	Mosquito Abundance	Vector Index current week	Vector Index		X
		Vector Index one week before	VIlag1		X
		Vector Index two weeks before	VIlag2		X
		Vector Index three weeks before	VIlag3		X
		Vector Index four weeks before	VIlag4		X
		Light and gravid trap collection mean current week	Trap_Mean		X
		Light and gravid trap collection mean one week before	Trap_Meanlag1		X
		Light and gravid trap collection mean two weeks before	Trap_Meanlag2		X
		Light and gravid trap collection mean three weeks before	Trap_Meanlag3		X
		Light and gravid trap collection mean four weeks before	Trap_Meanlag4		X
	Mosquito Biting Rates (HLC)	Mosquitoes per visit	mosquitoes per visit		X
		Culex spp. per visit	Cx per visit		X
Weather	Temperature	Average temperature current week	tempc		X
		Average temperature of one week before	templag1	X	X
		Average temperature of two weeks before	templag2	X	X
		Average temperature of three weeks before	templag3	X	X
		Average temperature of four weeks before	templag4	X	X
		Mean January temperature	Jantemp	X	X
	Precipitation	Average precipitation current week	preci		X
		Average precipitation of one week before	precilag1	X	X
		Average precipitation of two weeks before	precilag2	X	X
		Average precipitation of three weeks before	precilag3	X	X
		Average precipitation of four weeks before	precilag4	X	X

Table 2.1. (continued)

Designation		Covariate Information: Description	Notation	Cook/DuPage Model	Ultra-fine-scale Model
Anthropogenic	Socio-demographic	Percentage of White population	whitepct	X	X
		Percentage of African American population	blackpct	X	X
		Percentage of Asian population	asianpct	X	X
		Percentage of Hispanic population	hispanicpct	X	X
		Median household income	Income	X	X
		Percentage of housing constructed before WWII	hpctpreww	X	X
		Percentage of housing constructed post WWII (1945-1969)	hpctpostww	X	X
		Percentage of housing constructed from 1970-1989	hpct7089	X	X
		Percentage of housing constructed in 1990 or later	hpctpost90	X	X
	Land change & manipulation	Catch basin density	CB		X
		Total area of building structures	bldg.._footprint_area_total		X
		Average area of building structures	bldg._footprint_area_avg		X
		Total perimeter of building structures	Building_Footprint_peri_total		X
		Average perimeter of building structures	Building_Footprint_peri_avg		X
		Total area of residential lot	Residential_lot_area_total		X
		Average area of residential lot	Residential_lot_area_avg		X
		Total perimeter of residential lot	Residential_lot_peri_total		X
		Average perimeter of residential lot	Residential_lot_peri_avg		X
		Ratio of total building area by total lot area	total_bldg._area/total_lot_area		X
		Ratio of average building area by average lot area	avg_bldg._area/avg_lot_area		X
		Ratio of total building perimeter by total lot area	total_bldg._peri/total_lot_area		X
		Ratio of average building perimeter by average lot area	avg_bldg._peri/avg_lot_area		X
		Number of buildings	buildings		X
		Building density per mi.^{2}	bldg._density		X
		Number of residents per building	persons_per_bldg.		X
	Human population	Total human population	totpop	X	X
		Mean light pollution	lightpol		X
	Activity Observations	Senior Citizen Observations per visit	Senior_obs per visit		X
		Adults Observations per visist	Adults_obs per visit		X
		Children Observations per visit	Child_obs per visit		X
		Male Observations per visit	Male_obs per visit		X
		Female Observations per visit	Female_obs per visit		X
		Total Observations per visit	Total_obs per visit		X
Other		Year	yr	X	X
		Hexagon Designation	hexid	X	X
			Total Covariates Evaluated	**40**	**82**

Table 2.2. Description of selected hexagons (n=55) by residual categorization (PR1, PR0, NR1, LR1, LR0) within the Northwest Mosquito Abatement District (NWMAD). Descriptions of residual categorizations are as follows: PR = positive residual (underprediction, residuals ≥ 1.0), NR = negative residual (great overprediction, residuals ≤ -1.0), LR = low residuals (prediction close to actual, residuals -1.0 < X < 1.0). Values following the residual categorizations designated as: 1 = at least one human WNV case between 2005-2016; 0 = no human WNV cases between 2005-2016.

HexID	Cases	Residual	Category	Field Season
4349	0	-1.001	NR0	1
4806	0	-1.001	NR0	1
4241	0	-1	LR0	1
4854	0	-1	LR0	1
5250	0	-1	LR0	1
4250	1	-0.271	LR1	1
4471	2	0.877	PR1	1
4183	2	0.902	PR1	1
4984	1	0.912	PR1	1
5188	0	1.134	PR0	1
4597	1	1.531	PR1	1

HexID	Cases	Residual	Category	Field Season
4014	0	-1.001	NR0	Both
4082	0	-1.001	NR0	Both
4217	0	-1.001	NR0	Both
4415	0	-1.001	NR0	Both
4467	0	-1	LR0	Both
5199	0	-1	LR0	Both
5286	0	-1	LR0	Both
4313	3	0.033	LR1	Both
5235	1	0.055	LR1	Both
4609	2	0.399	LR1	Both
4637	1	1.027	PR1	Both
4332	1	1.279	PR1	Both
4335	1	1.767	PR1	Both
4676	1	1.838	PR1	Both
4449	3	1.841	PR1	Both
5239	0	2.198	PR0	Both
4743	2	4.881	PR1	Both
5181	1	17.057	PR1	Both
4617	0	18.013	PR0	Both

HexID	Cases	Residual	Category	Field Season
4242	0	-1.001	NR0	2
4614	0	-1.001	NR0	2
4181	0	-1	NR0	2
4381	0	-1	NR0	2
4382	0	-1	NR0	2
4578	0	-1	NR0	2
4923	0	-1	NR0	2
5185	0	-1	NR0	2
5262	0	-1	NR0	2
5234	1	-0.338	LR1	2
4070	1	-0.32	LR1	2
4952	1	-0.193	LR1	2
4073	1	-0.038	LR1	2
4678	1	-0.014	LR1	2
4135	1	-0.002	LR1	2
4104	1	0	LR1	2
4334	1	0.024	LR1	2
4243	2	0.071	LR1	2
5126	1	0.11	LR1	2
4065	1	0.181	LR1	2
5265	1	0.221	LR1	2
4098	1	5.557	PR1	2
4636	1	6.481	PR1	2
4346	1	6.967	PR1	2

Table 2.3. Model fit comparisons of the UFS hexagons, applying **(A)** newly acquired data (excluding HLC and human observations, covariate set 2), or **(B)** only the covariates made available to the previously published Cook/DuPage model (covariate set 4). Each model outcome was assessed using logistic (presence/absence WNV human illness case) and generalized linear (WNV case rates, controlling for human population) methods.

A.

Regression Method[a]	Model Environment	Included Covariates	df	p-value	ROC[b]	BIC	ΔBIC
1. Logistic	1. MIR & Mosquito Abundance ($E_1C_2O_1$)	- tempc - preci + templag1 + templag3* - precilag1* - precilag3 - precilag4 + mir mean - mir diff + mirlag1 + mirlag2 + mirlag3** + mirlag4* + totpop + blackpct + asianpct + dmipct - dhipct + ccpct - hpctpreww - hpctpostww - hpct7089 + abund - abundlag1 + abundlag2 + abundlag3 + abundlag4 + bldg. footprint area avg. - bldg. footprint peri avg. + resi lot peri total - resi lot peri avg. - avg. bldg. area:avg. lot area - total bldg. peri:total lot area	34	<0.0001	0.89	768.7	128.3
	2. Vector Index ($E_2C_2O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag4 + totpop + blackpct + asianpct + hispanicpct - Income + dospct - dmipct - hpctpreww - hpct7089 + CB + avg. bldg. area:avg. lot area*** + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4	23	<0.0001	0.86	661.1	20.7
	3. Best-Fit ($E_3C_2O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag3 + mirlag3** + totpop - whitepct + blackpct - Income + dospct + dmipct - hpctpreww - hpct7089 - abundlag1 - bldg. footprint area total + avg. bldg. area:avg. lot area** + VIlag4*	21	<0.0001	0.86	640.4	0
	4. Global ($E_4C_2O_1$)	- tempc* - preci* - yr + templag1* + templag2 + templag3 - templag4 – precilag1** - precilag2 - precilag3 – precilag4* + MIRmean - MIRdiff + mirlag1 + mirlag2 + mirlag3* + mirlag4 + totpop - whitepct - blackpct - asianpct - hispanicpct + income - owpct - dospct - dlipct - dmipct - dhipct - blpct - dfpct - mfpct - shrubpct - glandpct - pasturepct - wwpct - dtotpct - ftotpct + wtotpct - Jantemp + hpctpreww + hpctpostww + hpct7089 + hpctpost90 - abund - abundlag1 - abundlag2 + abundlag3 + abundlag4 + CB - bldg. footprint total area + bldg. footprint area avg. - bldg. footprint peri total - bldg. footprint peri avg. + resi lot area total + resi lot area avg. + resi lot peri total - resi lot peri avg. + total bldg. area:total lot area + VI - VIlag1 - VIlag2 - VIlag3 - VIlag4 + Light pol + NDVI	65	0.0009	0.97	833.5	193.1
2. Linear	1. MIR & Mosquito Abundance ($E_1C_2O_2$)	mir mean + mir diff + mirlag1 + mirlag2 + mirlag3 + mirlag4* + blpct** + abund - abundlag1 - abundlag2 + abundlag3 + abundlag4* + bldg. footprint area avg.***	13	<0.0001	N/A	-182037	3358
	2. Vector Index ($E_2C_2O_2$)	bldg. footprint area avg.*** + avg. bldg. area:avg. lot area*** + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4***	7	<0.0001		-185373	22
	3. Best-Fit ($E_3C_2O_2$)	- mirlag4 + bldg. footprint area avg.***+ avg. bldg. area:avg. lot area** + VIlag4***	4	<0.0001		-185395	0
	4. Global ($E_4C_2O_2$)	- tempc - preci + yr - templag1 + templag2 + templag3 + templag4 - precilag1 + precilag2 - precilag3 - precilag4 + MIRmean - MIRdiff + mirlag1 - mirlag2 + mirlag3 - mirlag4** - totpop - whitepct + blackpct + asianpct + hispanicpct + income + owpct - dospct - dlipct - dmipct - dhipct - blpct - dfpct - mfpct + shrubpct - glandpct + pasturepct + wwpct - dtotpct - ftotpct + wtotpct + Jantemp - hpctpreww - hpctpostww - hpct7089 - hpctpost90 - abund - abundlag1 - abundlag2 + abundlag3 + abundlag4 + CB - bldg. footprint total area - bldg. footprint area avg. + bldg. footprint peri total + bldg. footprint peri avg. - resi lot area total - resi lot area avg. + resi lot peri total - resi lot peri avg. + total bldg. area:total lot area + - avg. bldg. area:avg. lot area - # blgds - VI - VIlag1 - VIlag2 - VIlag3 + VIlag4*** + Light pol + NDVI	68	0.0044		-94558.2	90836.8

Table 2.3. (continued)

B.

Regression Method	Model Environment	Included Covariates	df	p-value	ROC	BIC	ΔBIC
1. Logistic	0. Cook/DuPage[c] ($E_0C_4O_1$)	- Yr - templag2 + templag3* + templag4* - Jantemp + mirlag1 + mirlag2 + mirlag3 + mirlag4* + totpop - owpct - dlipct - dfpct - glandpct + hpctpost90	15	<0.0001	0.85	632.3	56.1
	1. MIR & Mosquito Abundance ($E_1C_4O_1$)	dhipct - mfpct - wwpct + mirlag1 + mirlag2 + mirlag3* + mirlag4 + templag3 + templag4* - precilag1* - precilag2 – precilag4* + asianpct** + totpop + mir mean - mir diff + abund – abundlag1* + abundlag2 + abundlag3 + abundlag4	21	<0.0001	0.88	653.3	77.1
	2. Vector Index ($E_2C_4O_1$)	dhipct - mfpct - wwpct + templag3* + templag4* - precilag1* - precilag2 – precilag4* + asianpct + VIlag1 + VIlag2 + VIlag3* + VIlag4	13	<0.0001	0.86	576.2	0
	3. Best-fit 55 hex ($E_3C_4O_1$)	dhipct - mfpct - wwpct + mirlag3* + mirlag4 + templag3* + templag4** - precilag1* - precilag2 – precilag4* + asianpct** + totpop	12	<0.0001	0.89	580.8	4.6
	4. Global ($E_4C_4O_1$)	Yr - dospct - dlipct - dmipct - dhipct - dfpct - mfpct + blpct - shrubpct - glandpct - pasturepct - ccpct - wwpct - owpct + mirlag1 + mirlag2 + mirlag3* + mirlag4 - templag1 - templag2 + templag3* + templag4* - precilag1* - precilag2 - precilag3 – precilag4* + whitepct - blackpct - asianpct + hispanicpct - Income - totpop + Jantemp	33	<0.0001	0.90	792.0	215.8
2. Linear	0. Cook/DuPage ($E_0C_4O_2$)	Yr - templag2 + templag3 + templag4 + Jantemp + mirlag1 + mirlag2 + mirlag3* + mirlag4** - totpop* - owpct* - dlipct – dfpct* - glandpct + hpctpost90	15	<0.0001	N/A[b]	-227354	90
	1. MIR & Mosquito Abundance ($E_1C_4O_2$)	Dmipct** + blpct** + mirlag1 + mirlag2 + mirlag3 + mirlag4* + templag3 - precilag1 - precilag4 – totpop* - mir mean + mir diff - abund - abundlag1 + abundlag2 - abundlag3 + abundlag4*	17	<0.0001		-182001	45443
	2. Vector Index ($E_2C_4O_2$)	dlipct + dmipct*** + templag3* - totpop** - VI - VIlag1 + VIlag2 + VIlag3* + VIlag4***	9	<0.0001		-185347	42097
	3. Best-fit 55 hex ($E_3C_4O_2$)	Dmipct** + blpct** + mirlag3** + mirlag4** + templag3* - precilag1 - precilag4 – totpop*	8	<0.0001		-227444	0
	4. Global ($E_4C_4O_2$)	Yr - dospct - dlipct - dmipct - dhipct - dfpct - mfpct + blpct - shrubpct - glandpct - pasturepct - ccpct - wwpct - owpct + mirlag1 + mirlag2 + mirlag3* + mirlag4* - templag1 - templag2 + templag3 + templag4 - precilag1 - precilag2 - precilag3 - precilag4 + whitepct - blackpct - asianpct + hispanicpct - Income – totpop* + Jantemp	33	<0.0001		-227199	245

[a]Logistic regression outcome = human WNV presence/absence per hexagon, per week; GLM outcome = WNV human case rate (per hexagon, per week).
[b]ROC applies to only logistic regression.
[c]As the final selected model in the Original Cook/DuPage paper (2019), this model environment was assessed only for the comparison to the Cook/DuPage models for this study and not applied to the UFS model. The original model covariates, eftpct and ehwpct, have 0 observations among the selected 55 hexagons and were removed.

Table 2.4. Model fit comparisons of the UFS hexagons, using best-fit models with additional human landing catch and human activity observations to incorporate added human risk. Human risk covariates were added to the UFS model by **(A)** best-fit integration (covariate set 1) and **(B)** force-fitting (covariate set 3).

A.

Regression Method	Model Environment	Included Covariates	df	p-value	AUC	BIC	ΔBIC
1. Logistic	1.MIR & Mosquito Abundance ($E_1C_1O_1$)	- tempc - preci + templag1 - templag2 + templag3* - precilag1* - precilag3 - precilag4 + mirmean - mirdiff + mirlag1 + mirlag2 + mirlag3** + mirlag4 - whitepct + blackpct + dospct - dmipct - hpctpreww - hpct7089 + abund - abundlag1 + abundlag2 + abundlag3 + abundlag4 + bldg. footprint area total - total bldg. area:total lot area + male obs per visit + female obs per visit + *Culex* per visit	30	<0.0001	0.87	742.5	107.9
	2. Vector Index ($E_2C_1O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag4 - whitepct + blackpct + dmipct - hpctpreww - hpctpostww - hpct7089 + male obs per visit + female obs per visit + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4	26	<0.0001	0.86	692.7	58.1
	3. Best-Fit ($E_3C_1O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag4 + mirlag3** + blackpct - dospct + dmipct - hpctpreww - hpct7089 + abund + Light pol + male obs per visit + female obs per visit + *Culex* per visit + VIlag4	19	<0.0001	0.84	634.6	0
	4. Global[a] ($E_4C_1O_1$)	- tempc* - preci* - yr + templag1* + templag2 + templag3 - templag4 - precilag1** - precilag2 - precilag3 - precilag4* + mirmean - mirdiff + mirlag1 + mirlag2 + mirlag3* + mirlag4 - totpop + whitepct + blackpct + asianpct + hispanicpct + income - owpct - dospct - dlipct - dmipct - dhipct - blpct - dfpct - mfpct - shrubpct - glandpct - pasturepct + wwpct - dtotpct - ftotpct + wtotpct - Jantemp + hpctpreww + hpctpostww + hpct7089 + hpctpost90 - abund - abundlag1 - abundlag2 - abundlag3 + abundlag4 + CB - bldg. footprint total area - bldg. footprint area avg. + bldg. footprint peri total + bldg. footprint peri avg. - resi lot area total - resi lot area avg. - resi lot peri total + resi lot peri avg. - total bldg. area:total lot area - # blgds + VI - VIlag1 - VIlag2 - VIlag3 - VIlag4 + Light pol + NDVI + mosquitoes per visit	65	0.0009	0.97	772.8	138.2
2. Linear	1. MIR & Mosquito Abundance ($E_1C_1O_2$)	mir mean + mir diff + mirlag1 + mirlag2 + mirlag3 + mirlag4* + blpct** + abund - abundlag1 - abundlag2 + abundlag3 + abundlag4* + bldg. footprint area avg.***	13	<0.0001	N/A	-182037	3352
	2. Vector Index ($E_2C_1O_2$)	templag3* + bldg. footprint area avg.*** - resi lot peri avg. + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4***	8	<0.0001		-185362	27
	3. Best-Fit ($E_3C_1O_2$)	- mirlag4 + bldg. footprint area avg.***- resi lot peri avg. + VIlag4***	4	<0.0001		-185389	0
	4. Global[b] ($E_4C_1O_2$)	- tempc - preci + yr - templag1 + templag2 + templag3 + templag4 - precilag1 + precilag2 - precilag3 - precilag4 + mirmean - mirdiff + mirlag1 - mirlag2 + mirlag3 – mirlag4** - totpop + whitepct + blackpct + asianpct + hispanicpct + income - owpct - dospct - dlipct - dmipct - dhipct - blpct - dfpct - mfpct - shrubpct - glandpct - pasturepct + wwpct - dtotpct - ftotpct + wtotpct + Jantemp - hpctpreww - hpctpostww - hpct7089 - hpctpost90 - abund - abundlag1 - abundlag2 - abundlag3 + abundlag4 + CB - bldg. footprint total area - bldg. footprint area avg. + bldg. footprint peri total + bldg. footprint peri avg. - resi lot area total - resi lot area avg. + resi lot peri total - resi lot peri avg. - total bldg. area:total lot area - avg. bldg. area:avg. lot area - # blgds - VI - VIlag1 - VIlag2 - VIlag3 + VIlag4*** + Light pol + NDVI + senior obs per visit	64	0.0022		-94581.4	90797.6

Table 2.4. (continued)

B.

Regression Method	Model Environment	Included Covariates	df	p-value	ROC[b]	BIC	ΔBIC
1. Logistic	0. Cook/DuPage ($E_0C_3O_1$)	- Yr - templag2 + templag3* + templag4* - Jantemp + mirlag1 + mirlag2 + mirlag3* + mirlag4* - totpop - owpct - dlipct - dfpct - glandpct + hpctpost90 - senior obs per visit - adult obs per visit + child obs per visit + male obs per visit - mosquitoes per visit + *Culex* per visit	21	<0.0001	0.86	683.4	10.7
	1. MIR & Mosquito Abundance[c] ($E_1C_3O_1$)	- tempc - preci + templag1 + templag2 - precilag1* - precilag4* - mirmean* - mirdiff + mirlag1 + mirlag2 + mirlag3** - mirlag4* - totpop + blackpct + dospct + hpctpostww + hpct7089 + hpctpost90* + abund - abundlag1 + abundlag2 + abundlag3 + abundlag4 + bldg. footprint peri avg. + resi lot area total + avg. bldg. peri:avg. lot area** + senior obs per visit + adult obs per visit + child obs per visit - male obs per visit + mosquitoes per visit + *Culex* per visit	32	<0.0001	0.87	757.7	61.1
	2. Vector Index ($E_2C_3O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag4 - whitepct + blackpct - dospct + dmipct + hpctpostww + hpct7089 + hpctpost90 + resi lot area total - senior obs per visit + adult obs per visit + child obs per visit - male obs per visit - mosquitoes per visit + *Culex* per visit + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4	26	<0.0001	0.85	696.6	23.9
	3. Best-fit 55 hex[d] ($E_3C_3O_1$)	- tempc - preci + templag1 + templag2 + templag3** - precilag1* - precilag4 - whitepct + blackpct - dospct + dmipct + hpctpostww + hpct7089 + resi lot area total - senior obs per visit + adult obs per visit + child obs per visit - male obs per visit - mosquitoes per visit + *Culex* per visit + VIlag4 + mirlag3** - abund	23	<0.0001	0.84	672.7	0
	4. Global ($E_4C_3O_1$)	Model Failed to Converge	N/A	N/A	N/A	N/A	N/A
2. Linear	0. Cook/DuPage[e] ($E_0C_3O_2$)	Yr - templag2 + templag3 + templag4 + Jantemp + mirlag1 + mirlag2 + mirlag3* + mirlag4** - totpop* - owpct* - dlipct – dfpct* - glandpct + hpctpost90 - senior obs per visit + adult obs per visit + child obs per visit - male obs per visit + mosquitoes per visit - *Culex* per visit	21	<0.0001	N/A	-227300	0
	1. MIR & Mosquito Abundance[e] ($E_1C_3O_2$)	blpct** + bldg. footprint area avg.*** - senior obs per visit + adult obs per visit + child obs per visit - male obs per visit - mosquitoes per visit + *Culex* per visit + mirmean + mirdiff + mirlag1 + mirlag2 + mirlag3 + mirlag4* + abund - abundlag1 - abundlag2 - abundlag3 + abundlag4*	19	<0.0001		-181982	45318
	2. Vector Index[f] ($E_2C_3O_2$)	bldg. footprint area avg.*** + adult obs per visit + child obs per visit + female obs per visit - total obs per visit - mosquitoes per visit + *Culex* per visit + VI - VIlag1 + VIlag2 + VIlag3* + VIlag4***	12	<0.0001		-185322	41978
	3. Best-fit 55 hex[g] ($E_3C_3O_2$)	bldg. footprint area avg.*** - senior obs per visit - adult obs per visit + child obs per visit + female obs per visit - mosquitoes per visit + *Culex* per visit + VIlag4*** - mirlag4	9	<0.0001		-185344	41956
	4. Global[g] ($E_4C_3O_2$)	- preci + Yr - templag1 + templag2 + templag3 + templag4 - precilag1 + precilag2 - precilag3 - precilag4 - totpop + whitepct + blackpct + asianpct + hispanicpct - Income - owpct - dospct - dlipct - dmipct - dhipct + blpct - dfpct - mfpct - shrubpct + glandpct - pasturepct* - wwpct - dtotpct - ftotpct + wtotpct + Jantemp - hpctprewww - hpct7089 - hpctpost90 + CB - bldg. footprint area total - bldg. footprint area avg. + bldg. footprint peri total + lightpol + NDVI + senior obs per visit - adult obs per visit - child obs per visit + female obs per visit - mosquitoes per visit + *Culex* per visit - VI - VIlag1 - VIlag2 - VIlag3 + VIlag4*** + mirmean - mirdiff + mirlag1 - mirlag2 + mirlag3 - mirlag4** - abund - abundlag1 - abundlag2 - abundlag3 + abundlag4	63	0.0013		-94601.5	132699

[a]avg. bldg. peri:avg. lot area, # bldg.s, Observations per visit: senior, adult, child, male, female, total human and Culex spp. were biased and/or zeroed and not assessed.
[b]Observations per minute: adult, child, male, female, total human, mosquito, and Culex spp. were biased and/or zeroed and not assessed.
[e]Observations per minute: female and total human were biased and/or zeroed and not assessed.
[f]Observations per minute: male and senior were biased and/or zeroed and not assessed.
[g]Observations per minute: male and total human were biased and/or zeroed and not assessed.

Table 2.5. Detailed assessment of each model evaluated in this study. Overall model strength was determined by BIC value (by linear and logistic regression types), with the following characteristics denoted as follows: Cumulative Significance Total, sum of each variable score, denoted as: $p<0.001 = 4$, $p<0.01 = 3$, $p<0.05 = 2$, included in model = 1; # of Significant Covariates = summation of included covariates with p-value<0.05; DF = degrees of freedom denoted in model; BIC value = overall model rank (best model = 1, worst model = 14)/14 for each logistic and linear model group, respectively.

Model[a]	Cumulative Significance Total	# of Significant Covariates	DF (lower is better)	BIC value	BIC value (lower is better) rank	Regression Type
E0C3O2	6	5	21	-227300	3	Linear
E0C4O2	6	5	15	-227354	2	
E1C1O2	7	4	13	-182037	11	
E1C2O2	7	4	13	-182037	12	
E1C3O2	7	4	19	-181982	14	
E1C4O2	7	5	17	-182001	13	
E2C1O2	8	4	8	-185362	7	
E2C2O2	10	4	7	-185373	6	
E2C3O2	7	3	12	-185322	10	
E2C4O2	10	5	9	-185347	8	
E3C1O2	6	2	4	-185389	5	
E3C2O2	8	3	4	-185395	4	
E3C3O2	6	2	9	-185344	9	
E3C4O2	10	6	8	-227444	1	
E4C1O2 E4C2O2 E4C3O2 E4C4O2	Global Models Excluded					
E0C3O1	4	4	21	683.40	9	Logistic
E0C4O1	3	3	15	632.30	3	
E1C1O1	4	3	30	742.50	12	
E1C2O1	5	4	34	768.70	14	
E1C3O1	9	7	32	757.70	13	
E1C4O1	7	6	21	653.30	6	
E2C1O1	4	3	26	692.70	10	
E2C2O1	7	4	23	661.10	7	
E2C3O1	4	3	26	696.60	11	
E2C4O1	5	5	14	576.20	1	
E3C1O1	5	3	19	634.60	4	
E3C2O1	8	5	21	640.40	5	
E3C3O1	5	3	23	672.70	8	
E3C4O1	8	6	12	580.80	2	
E4C1O1 E4C2O1 E4C3O1 E4C4O1	Global Models Excluded					

[a]All global models were excluded from analysis as they were all overfit and statistically biased

FIGURE 2.1. The UFS study area, contained within the Northwest Mosquito Abatement District, shown in relation to Cook & DuPage Counties. Overlaid are 1-km diameter hexagons, the observational units used in this study. Northwest Mosquito Abatement District comprises 1,019 of the total 5,345 hexagons in all of Cook & DuPage Counties.

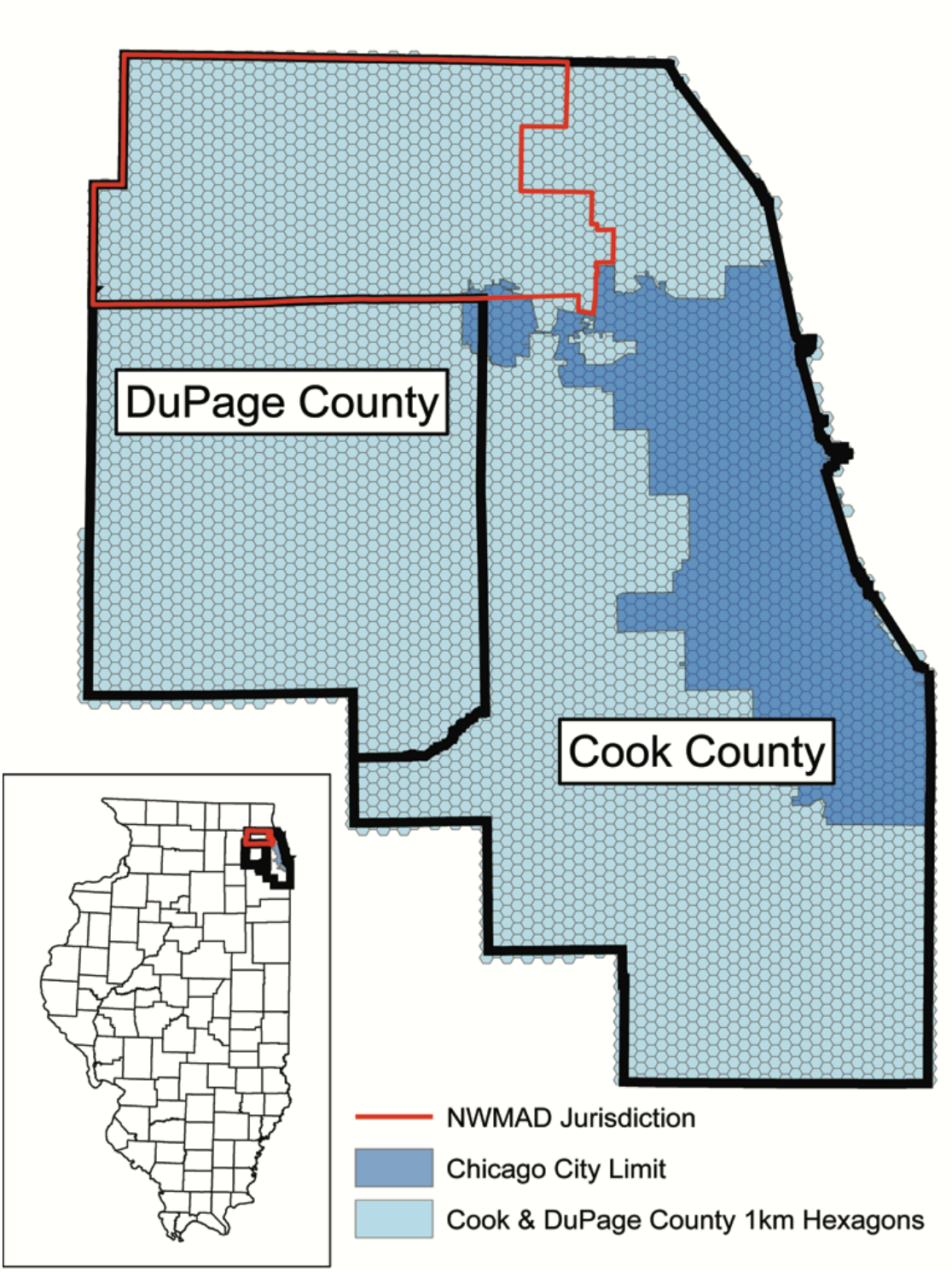


FIGURE 2.2. Location of the 55-hexagon study area within the Northwest Abatement District. Hexagons are labeled by field season visited for mosquito collections and human activity observations (color outline) and by total human cases from 2005-2016 (gray scale shaded interior).

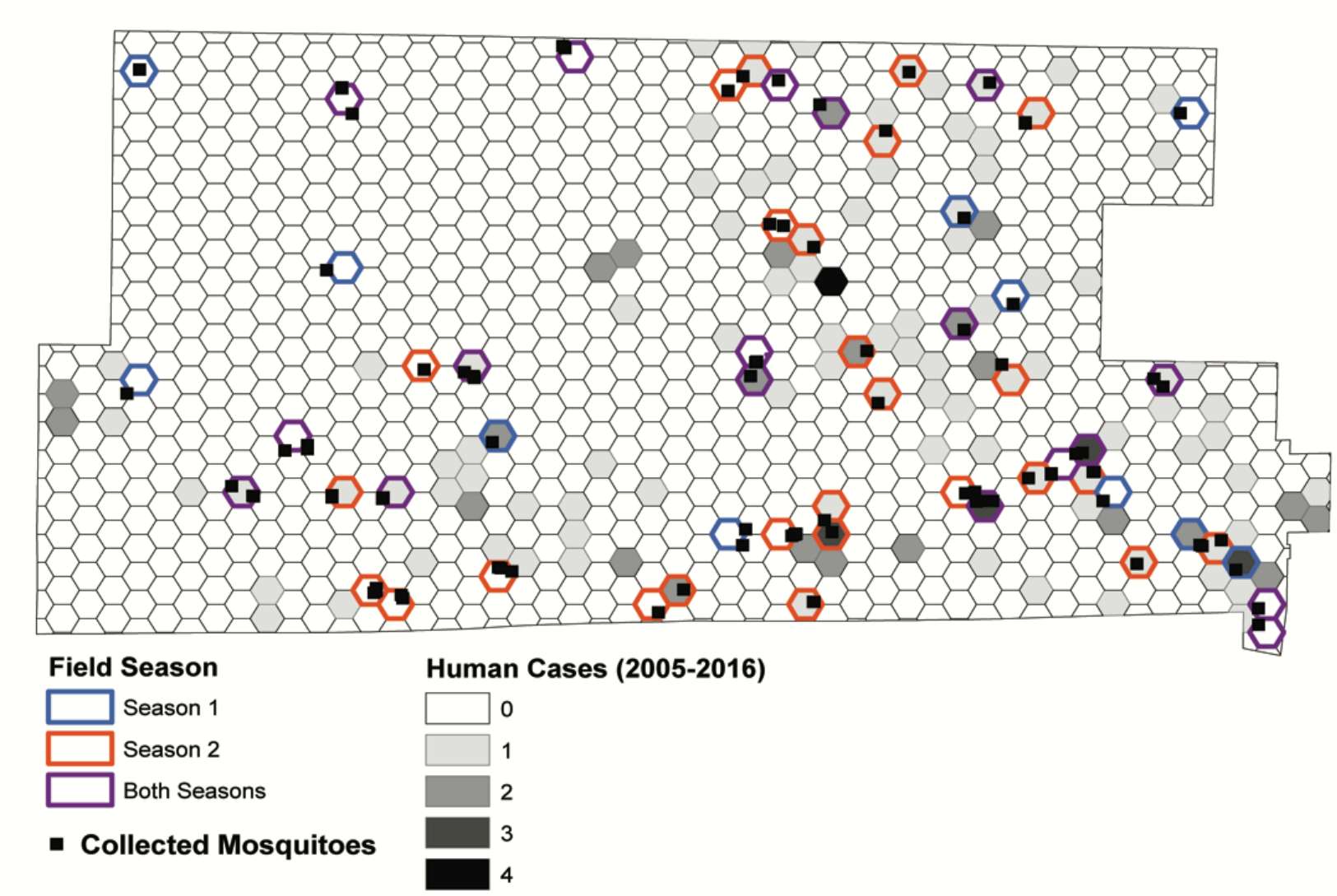


FIGURE 2.3. Mean performance of each of the 70 covariates used in the study (A). Covariates are listed in alphabetic order by data availability/workload to acquire score (1-4). The contribution of each covariate resulted in a net value performance for each linear and logistic model assessed in the study (B). Means for each outcome ($\bar{x}_{covariate}$ = 0.48; $\bar{x}_{linear}$ = -193406; $\bar{x}_{logistic}$ = 670.9) are designated by horizontal dashed lines. Details of scoring for each covariate and model are provided in Tables 2.5 & Appendix A, Table 1.

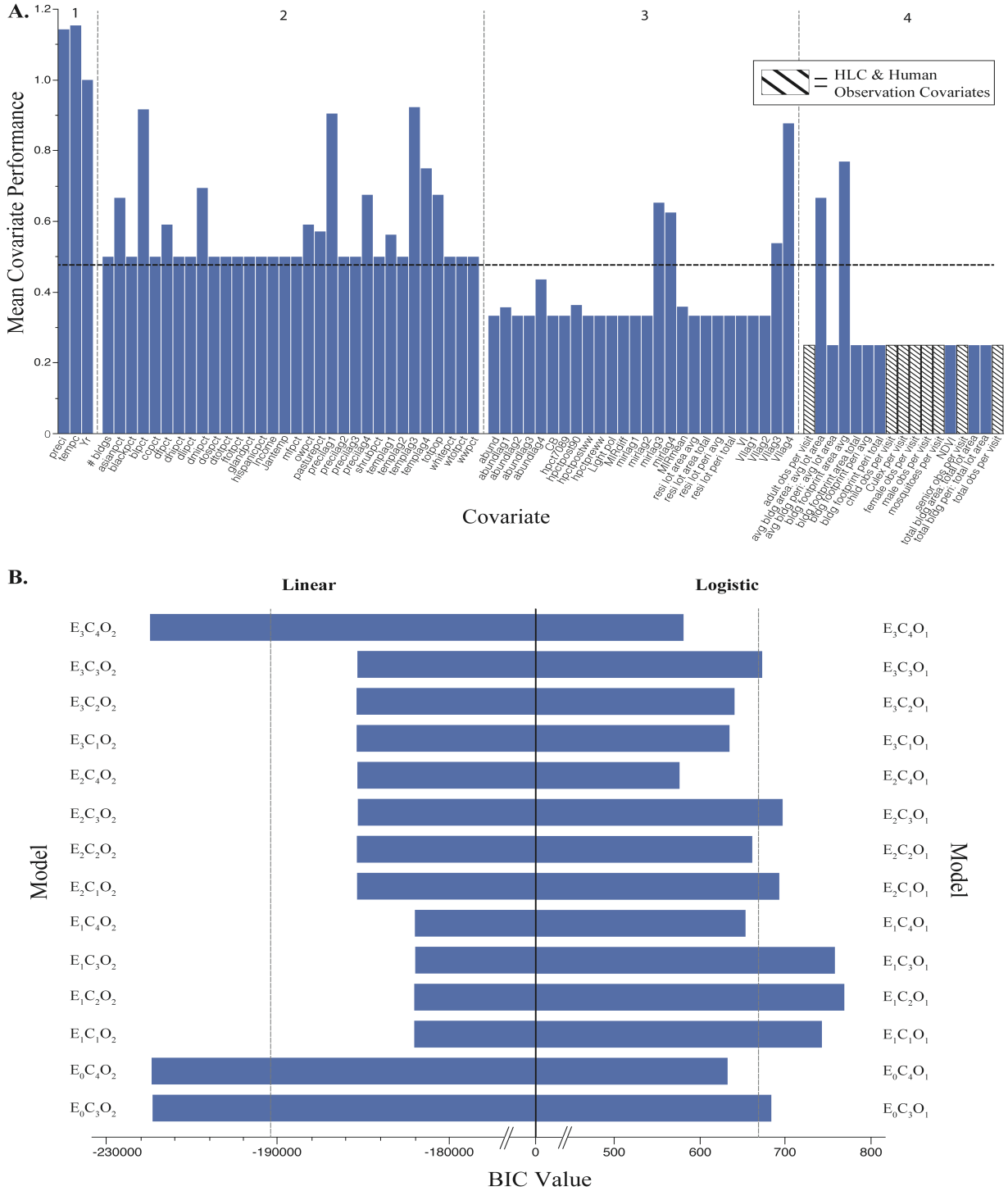


FIGURE 2.4. Relationship of hexagon type (LR = low residual, PR = positive residual, NR = large, negative residual; 0 = no human case, 1 = human case) by human observations per visit (A), mosquitoes collected per visit (B), and a product of the two former variables, nuisance factor and WNV added risk (C). Letters above each box and whisker plot designate significantly different groups by hexagon type, as calculated by Tukey's HSD.

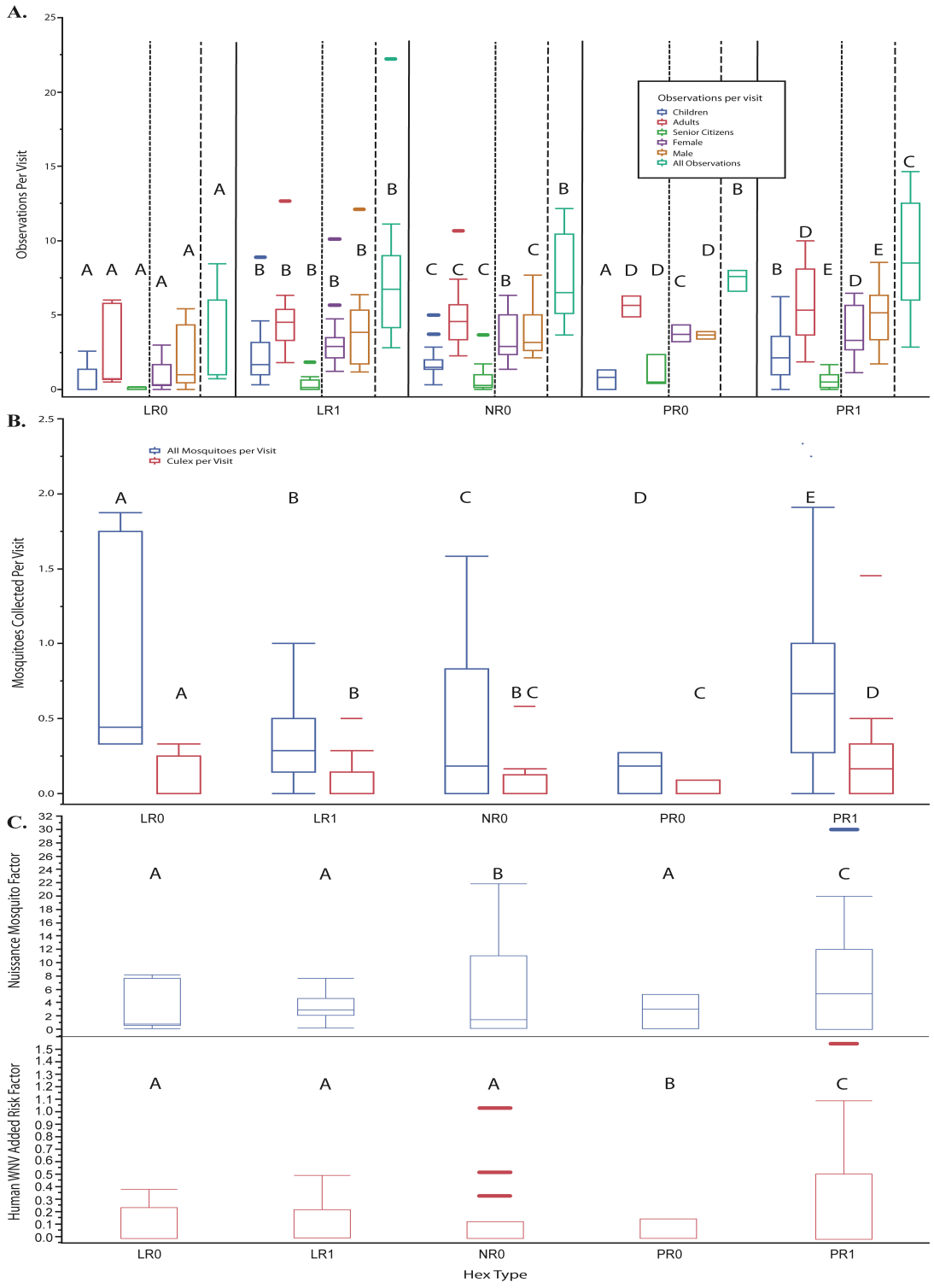
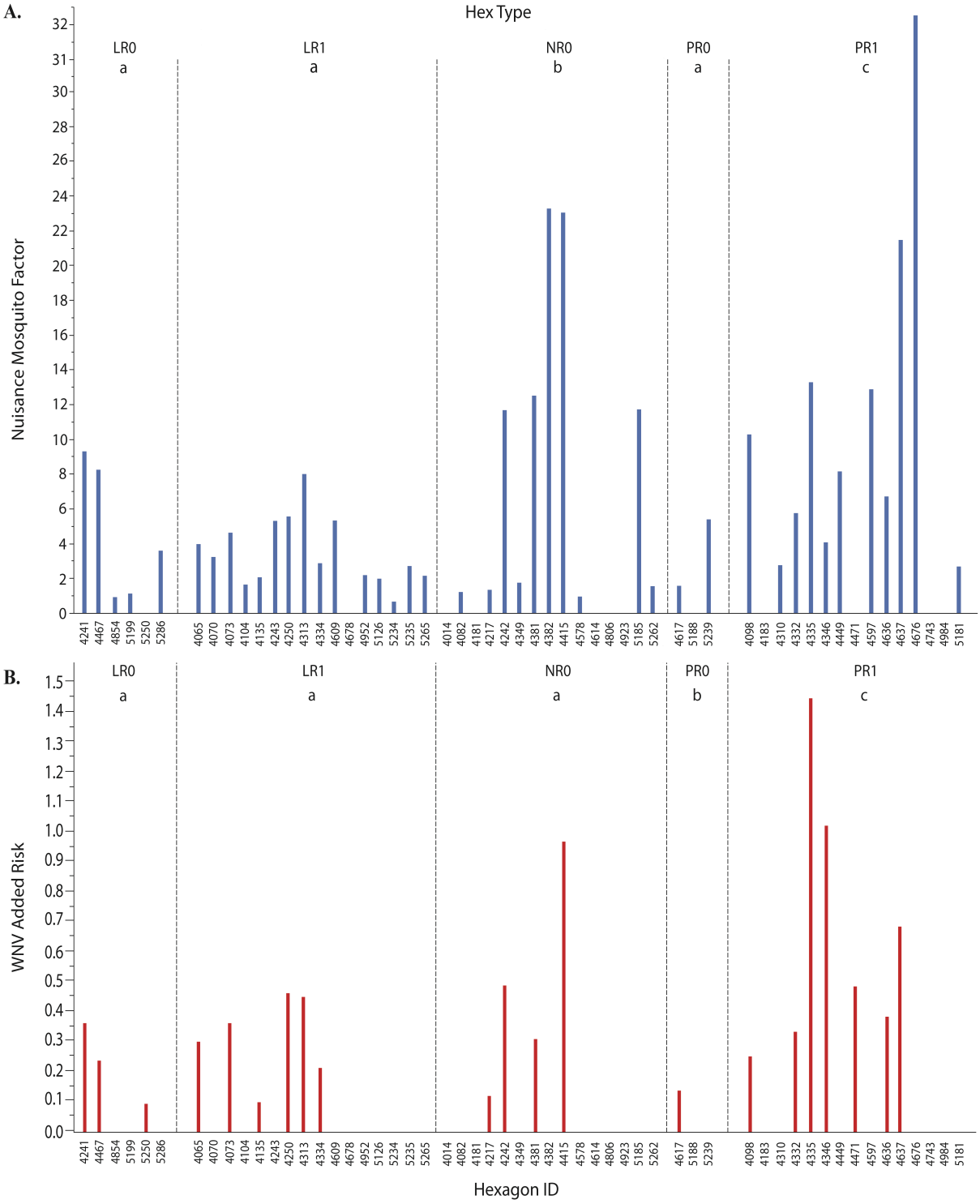


FIGURE 2.5. Relationship of each of the 55 UFS study hexagons (y-axis = unique identification number) by nuisance mosquito factor (A) and human WNV added risk (B). Letters above each box and whisker plot designate significantly different groups by hexagon type, as calculated by Tukey's HSD.



CHAPTER 3: EFFECTS OF SCALE ON MODELING WEST NILE VIRUS DISEASE RISK

3.1. ABSTRACT

Modeling vector-borne disease transmission is a difficult process, partly due to numerous cross-species interactions. These interactions are affected by a wide range of external forces associated with climate and land use. Altogether, these forces strongly influence the local and microscale dynamics of disease. West Nile virus (WNV) is the most important mosquito-borne disease in the United States. Vectored by *Culex* mosquitoes and maintained in the environment by avian hosts, the virus can spillover into humans and horses, sometimes causing severe neuroinvasive illness. Several studies have evaluated drivers of WNV disease risk, but nearly all have done so at coarse scales, and have reported mixed results of the effects of common explanatory variables. As a result, fine-scale relationships with common explanatory variables, especially those related to ecological processes, remain uncertain across varying spatial extents. Using an interdisciplinary approach and an ongoing 12-year study of the Chicago region, this study evaluated the factors explaining WNV disease risk at high spatiotemporal resolution, comparing human WNV model and covariate performance across three increasing spatial extents: ultra-fine, local, and county scales. Our results demonstrate that as spatial extent increased, model performance increased. Additionally, only 26.1% of the twenty-three assessed covariates were included in best-fit models of at least two scales. These results suggest that the mechanisms driving WNV ecology are scale-dependent and covariate importance increases as extent decreases. These tools may be particularly helpful for

public health, mosquito, and disease control personnel in predicting and preventing disease within local and fine-scale jurisdictions, before spillover occurs.[3]

3.2. INTRODUCTION

Vector-borne diseases (VBDs), notorious for their ubiquitous, yet apparent low ability to inflict disease to humans and animals for multiple years, have the ability to suddenly erupt into seasonal outbreaks (Skaff and Cheruvelil 2016). Abiotic forces, most notably those that are climatic (e.g. temperature, precipitation, humidity), directly mediate the lifecycles of arthropod vectors, many of which are among the first species in an ecosystem to respond to seasonal changes (Patz and Reisen 2001, Patz et al. 2003, 2005, Semenza and Menne 2009, Lafferty 2010). Medically important mosquitos, like *Aedes* spp. and *Culex* spp. for example, are affected differently by temperature and precipitation. The former thrives in wet and warm conditions, while the latter thrives in dry and hot conditions. Depending on the vector species, under ideal abiotic conditions, the forces can provide conditions optimal for rapid population growth and development, biting opportunities, and pathogen amplification (Reither 2001, Lafferty 2010).

These forces are often coupled with biotic factors that strongly influence local and microscale dynamics of disease (Cohen et al. 2016). For example, a given county may be predominantly wetland, a habitat conducive for mosquito breeding, but the true prevalence of mosquito abundance can differ drastically within particular, smaller regions. Impervious land (land which does not allow water to pass through) is commonplace throughout cities and towns, and may affect the flow and drainage of water, creating

[3] This chapter submitted as an article to The American Journal of Tropical Medicine and Hygiene. The original citation appears as follows: Uelmen, J.A., Irwin, P., Bartlett, D., Brown, W.M., Karki, S., Ruiz, M.O., Fraterrigo, J., Li, B., and Smith 2020. Effects of Scale on Modeling West Nile virus Disease Risk. *The American Journal of Tropical Medicine & Hygiene,* Under Review.

artificially induced inundated and/or semi-permanent flooding events. Additionally, expansions of cities, towns, and the connectivity of humans (e.g. roads), often create patches of natural habitat. These patches, or fragments, become disrupted pieces of the landscape, facilitating microscale differences in ecological systems. In the provided example, a predominant wetland county with human disturbances creates an increasingly complex, heterogeneous mixture of abiotic and biotic forces on mosquito vectors, directly affecting disease ecology that is measurable only at high resolutions, (i.e. very fine scales) (Hurlbert and Jetz 2007, Meentemeyer et al. 2012, Skaff and Cheruvelil 2016).

West Nile virus (WNV) is a mosquito-borne *Flavivirus* that infects a vast array of vertebrate hosts (Kramer et al., 2008). In North America, it is predominantly transmitted from mosquitoes of the *Culex* genus, and the predominant *Culex* vector species differs by region (Goddard et al. 2002). The enzootic cycle of WNV occurs when an infected mosquito takes a blood meal from an uninfected avian host, or when an uninfected mosquito takes a blood meal from an infected, and actively shedding, infected avian host (Kilpatrick et al. 2007, Hamer et al. 2009). The susceptibility and infectiousness of an avian host varies by species, but birds from the family *Corvidae* and *Turdidae* are considered the most important for amplifying the disease (Kilpatrick et al. 2007, Hamer et al. 2009, Loss et al. 2009). In the midwestern United States, the enzootic cycle of WNV is predominately maintained by the mosquito vectors *Culex pipiens*, *Culex restuans*, and *Culex salinarius* (Andreadis et al. 2001, Petersen 2001, Molaei et al. 2006), and occasionally can spillover to dead-end (hosts that are not capable of infecting subsequent biting mosquitos) human and equine hosts (Bowen and Nemeth 2007). The most susceptible human hosts are those that are elderly and/or immunosuppressed

(Granwehr et al. 2004, Colpitts et al. 2012). An estimated 75-80% of infected humans are asymptomatic, and of the remaining 20-25%, <1% (or about 1 in 150 infected humans) will experience severe neuroinvasive disease (CDC 2020a). Cook and DuPage Counties, encompassing the greater Chicago, IL (USA) metropolitan region, have been among the hardest hit with human WNV in the country (CDC 2020b). Since the introduction of WNV in 2002, hotspots for human WNV illness in Chicago have occurred in high human-density locations within Chicago city and in the northern and northwestern suburbs, infecting hundreds despite differences in race, socioeconomic status, and other key demographics (Karki et al. 2020).

Many studies have modeled WNV risk, often with mixed results of the effects and magnitudes of several commonly reported covariates (Ruiz et al. 2004, Allan et al. 2009, Kilpatrick and Pape 2013, Manore et al. 2014, Roiz et al. 2014, Rosà et al. 2014, Wimberly et al. 2014, Hahn et al. 2015, Giordano et al. 2017, Keyel et al. 2019). Far fewer studies have investigated the effects of important forces on WNV disease risk at local scales (<100 km^2) (Gibbs, S., Wimberly, Michael, Madden, Marguerite, Masour, J., Yabsley, Michael, Stallknecht 2006, Meentemeyer et al. 2012). While there is no correct single scale to measure the entirety of a given disease (Wiens 1989, Levin 1992), studies conducted at regional or landscape scales may overlook fine-scale processes affecting disease dynamics at local scales because heterogeneity within a landscape moderates the broad-scale consistency of such processes (Levin 1992, Lawton 2016, Barker 2019), and are often not generalizable even within the regions where the studies take place (Cohen et al. 2016). Focusing on covariate and model selection at a fine scale and then applying to

broader scales will provide a better perspective and understanding of the heterogeneity of spatiotemporal processes that influence WNV disease ecology (Meentemeyer et al. 2012).

Translating fine-scale ecological processes into digestible epidemiological analyses has been immensely difficult, requiring both the resources to gather data at high spatiotemporal resolution and the computational hardware and technology to process such large datasets (Meentemeyer et al. 2012, Bansal et al. 2016). These limitations often force researchers to choose between increasing extent and decreasing grain (or vice-versa) (Levin 1992, Mayer and Cameron 2003). However, in recent years, high performance computer technology has become readily available and affordable (Cohen et al. 2016). Combining elements of ecology, epidemiology, entomology, spatial statistics, we use an existing and ongoing 12-year WNV dataset of the Chicago region, to evaluate the drivers of WNV eco-epidemiology at a high spatial and temporal resolution (30m to 1 km spatial x 1 week temporal resolution) across relatively large spatial extents. Specifically, we compared human WNV model and covariate performance across three increasing extents: 1. Ultra-fine-scale (UFS) subset of 55 study sites, 2. Local scale of 1019 sites, 3. County scale of 5345 sites. We hypothesized that as spatial scale increased (and thus, complexity in heterogeneity), model performance decreased, the number of covariates increased, and the magnitude in effect of each covariate will decrease.

3.3. METHODS

This project was approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign, the Illinois Department of Public Health (IDPH), and the University of Illinois Biosafety Committee. Human case data were provided by IDPH without any personal identifying information.

3.3.1. Study sites

The observational unit by which all models were evaluated consisted of 1 km-wide (0.65 km^2) hexagon shaped polygons. Hexagons were used for two notable reasons: 1. They are the most complex regular polygon that can continuously fill a 2-dimensional plane without gaps or overlap in configuration (less loss in orientation), and 2. The shape index ($perimeter^2$/area) is more compact than most other shapes (e.g. square or rectangle), providing more accurate sampling (Birch et al. 2007). All models had the same spatial[4] and temporal grain - 0.65 km^2 (1 km-width) and 1 week (denoted by CDC epidemiological weeks, beginning on Sunday and ending on Saturday), respectively. These resolutions were the finest scale by which our research team could reliably collect data throughout our study sites, due to limitations in data availability and revisit times.

This study compared and evaluated the performance of each model and individual covariate across three study sites, varying in spatial extent: 1) 55 hexagons, 92.5 km^2 ("ultra-fine-scale", UFS); 2) 1019 hexagons, 605.2 km^2 (local scale); 3) 5345 hexagons, 3471.7 km^2 (county scale) (Figure 3.1).

The UFS model, consisting of several individual field sites within each location, is the finest known spatial extent to evaluate WNV at weekly temporal resolutions (Uelmen, J.A., Irwin, P., Brown, W., Karki, S., Ruiz, M.O., Li, B., Smith 2020). The 55 hexagons within the model have been selected across a spectrum of performance (ranging from extremely poor to extremely well-fit), based on strength of prediction, as indicated by the residual output from a previous Cook-DuPage WNV model (Karki et al. 2020). The decision to choose 55 hexagons is the maximum number of sites that our researchers

[4] The UFS model included field data that was collected at a maximum of 30 m resolution. For this study, that data was then aggregated into the respective hexagon by which it was collected.

could visit for 15-minutes each, weekly, over two field seasons (June – September 2018, 2019) while providing adequate spatial coverage as a subset of the Northwest Mosquito Abatement District (NWMAD).

The local scale is an area consisting of 1019 hexagons. This region is an enclave of Cook County, encompassing the jurisdiction of one of the Chicago area's four mosquito abatement agencies, the NWMAD. This local scale was chosen as study site because close collaboration from the NWMAD provided several advantages, including: 1) Access to high quality and well maintained longitudinal datasets; 2) Permission to study and use equipment, if needed; 3) Provision of important local information (e.g. areas that are flood or mosquito prone); 4) Work towards a common goal to better understand and improve upon the safety of public health as it pertains to local WNV dynamics.

The Cook-DuPage region, denoted as the county scale, comprises of 5345 hexagons and is the largest extent at which we evaluated human WNV illness. Despite its large two county extent, the spatiotemporal resolution is still favorable for evaluating local-regional effects on WNV dynamics (1 km spatial grain x 1 wk temporal resolution).

3.3.2. Model Parameters

3.3.2.1. Dependent Variable - Human Illness

Human WNV cases in Illinois were classified as either confirmed or probable, as reported to the IDPH by public health or licensed medical professionals (mandatory reporting of WNV cases is required in the state)[5]. We recognize that exposure to

[5] The case definition for a confirmed case of arboviral encephalitis in Illinois is a clinically compatible illness that is laboratory confirmed at a public health laboratory. The laboratory criteria are a fourfold or greater rise in serum antibody titer; or isolation of virus from, or demonstration of viral antigen in, tissue, blood, CSF or other body fluid; or specific IgM antibody in CSF. A probable case of arboviral encephalitis is a clinically compatible illness occurring during the season when arbovirus transmission is likely to occur

mosquito-borne disease occurs often and in many locations, and that confirming the moment an infected mosquito inoculates a human is nearly impossible. We assumed for this model that human cases were the result of exposure at their home addresses, and the latitude and longitude point locations of each human case were provided to the third decimal degree and aggregated to the hexagon level for analytical and display purposes. Human cases were converted into binary form (presence/absence of illness), controlling for human population, by week for each hexagon. Use of human case data was approved by the University of Illinois Institutional Review Board and the Illinois Department of Public Health.

3.3.2.2. Independent Variables

The total number of independent variables available for each model varied by scale. The county, local, and ultra-fine scales had 40, 59, and 82 variables available, respectively (Table 3.1). Specific details pertaining to each independent variable's processing and data source is available in Uelmen et al. (2020).

3.3.2.2.1. Abiotic Predictors

Abiotic independent variables consisted of environmental (2011 United States Geological Survey (USGS 2011) National Land Cover Database (NLCD)) and weather (daily mean temperature and precipitation were acquired from the PRISM Climate Group (Oregon State University)) data. Additional independent variables included catch basin (e.g. sewer) density and light pollution (Falchi, F., Cinzano, P., Duriscoe, D., Kyba, C. C. M., Elvidge, C. D., Baugh, K., Portnov, B., Rybnikova, N. A., Furgoni 2016b, 2016a) per

---

and with the following supportive serology: a stable (twofold or smaller change) elevated antibody titer to an arbovirus, e.g., > 320 by hemagglutination inhibition, > 128 by complement fixation (CF), > 256 by IF, > 160 by neutralization, or a positive serologic result by enzyme immunoassay (EIA) or MAC EL

hexagon. There were 29 variables available to both the UFS and local scale and 23 variables available to the county scale.

3.3.2.2.2. Biotic Predictors

All mosquito infection data was provided by the Illinois Department of Public Health (IDPH), the state agency responsible for collecting and maintaining standardized mosquito collection and testing data. Mosquito abundance data was provided by the NWMAD and was included in the models for the local and ultra-fine scales, but not the county scale. As such, vector index (VI), a factor of both mosquito infection and abundance, was calculated and made available for the local and ultra-fine scale models. The following weekly mosquito infection indices were calculated by the following equations:

$$\text{Minimum Infection Rate (MIR)} = \frac{\#\ positive\ mosquito\ pools}{total\ specimens\ tested} \times 1000,$$

where a mosquito pool consisted of up to 50 female *Culex* mosquitoes that were collected by the same trap;

$$\text{VI} = \sum_{i=Culex\ spp.\ (pooled)} \bar{N}_i \hat{P}_i,$$

where $\bar{N}_i$ = average abundance (number of mosquitoes per trap week) and $\hat{P}_i$= estimated MIR.

The mosquito infection indices were paired with each trap location that they were derived from, and interpolated across NWMAD via inverse distance weighting (IDW) in a geographic information system (GIS), Environmental Systems Research Institute's (ESRI) ArcGIS platform (version 10.5.1, ESRI 2011). The average mosquito infection values were extracted for each hexagon using the zonal statistics tool in ArcGIS.

Total population and racial composition (White, African American, Hispanic, and Asian) at the census block level were extracted from the 2010 U.S. Census and converted to a percentage for each hexagon. The 2015 American Community Survey provided block level age of housing and income data that was aggregated by hexagon. Normalized difference vegetation index (NDVI) were processed from Landsat 7 and 8 bands for early, mid, and late summer periods from EarthExplorer (USGS 2019) and averaged by hexagon.

3.3.3. Statistical methods

3.3.3.1. Model Performance and Comparisons

Candidate covariates for each model were first screened by univariate analysis (cut-off p-value ≤ 0.20, Appendix B, Table 1). Covariates that passed the initial univariate screening were then selected for the final model, a generalized linear regression personality with a logit link function, via forward selection, based on the lowest Bayesian information criterion (BIC) value achieved. Although assessed across the same temporal period (weekly, 2005-2016), each of the three models corresponding to the three scales varied across spatial extent, resulting in increasing availability of data. Traditional performance metrics (e.g. AIC and BIC) are valid only for evaluating likelihood estimates across models that use the same dataset. Therefore, this study evaluated the overall performance of each model by root mean square error (RMSE), an evaluation of the standard deviation in residual values across each observation to the line of best fit. Receiver operating characteristic (ROC) curves were used to visualize each model's overall performance, and Area Under the Curve (AUC) values were calculated as secondary model performance indicators. All predictors were evaluated for

multicollinearity and analyzed using the Regression and Fit Model features, respectively, in JMP 14.2.0 (SAS Institute Inc. Cary, NC, USA).

Human WNV illness risk maps (from 2005-2016), generated from the best-fit models at each scale, were created and compared for differences in magnitude in the probability of cumulative human cases. This process was conducted by first creating a raster layer for each included covariate per respective model. A raster layer visualizes data as a surface represented by a regular grid of pixels (each pixel represents 30 m). Using the raster calculator tool in ArcGIS, each included covariate's estimate was input with its respective raster in a logistic regression (equation 1), then transformed to create a probability map of human WNV cases (equation 2).

Equation 1: Model "C, L, or U" = $-1 * (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots)$

Equation 2: WNV Risk Model = $\frac{1}{1+e^{Model\ "C,L,or\ U"}}$

Using the raster calculator again, the difference in probability by pixel of the larger scale[6] raster (either local or ultra-fine scale) from the smaller scale raster (either county or local scale) was calculated (equation 3).

Equation 3: $\frac{1}{1-(WNV\ Risk\ Model\ C\ or\ L - WNV\ Risk\ Model\ L\ or\ U)}$,

where model C = best-fit model for county scale, model L = best-fit model for local scale, and model U = best-fit model for ultra-fine scale model.

Positive values indicate areas where the larger scale raster overestimated WNV risk while negative values indicate where the WNV risk was underestimated by the larger

[6] Large scale (synonymous with fine scale)= small extent, more detail; Small scale (synonymous with broad or coarse scale) = large extent, less detail

scale raster. To evaluate performance due to scale-dependency, best-fit models from the smaller scales were also applied to larger scales, comparing BIC and ROC values to their original best-fit values.

3.3.3.2. Covariate Performance and Comparisons

Individual covariate performance was evaluated within each respective scale's best-fit model using the leave-one-covariate-out (LOCO) procedure, as introduced by Lei et al. (2018). This method was chosen for the following reasons: 1. It is robust and not limited to linear models, 2. It emphasizes the importance of a variable in a model as it pertains to prediction, and 3. Any algorithm can be used to measure the importance of the covariate and is computationally flexible (Tibshirani et al. 2018). After each of the three scales' respective best-fit models were chosen, the performance of each covariate was analyzed by removing it from the model and finding the difference in RMSE. Covariates removed with a percent difference in RMSE value extending further from zero in the positive direction indicated their increasing importance to the model, while a percent difference in RMSE value extending further from zero in the negative direction indicated decreasing importance to the model.

3.4. RESULTS

3.4.1. Descriptive Statistics of Each Scale Area

Between 2005-2016, there were a total of 906 reported human cases in the Cook-DuPage study area. Within this region and time frame, the local scale study area consisted of 156 human cases (17.2%) with the UFS study contributing 46 human cases (5.1%).

3.4.2. Comparisons

3.4.2.1. Models

Best-fit models of county and local scales resulted in similarly small RMSE values (0.024872 and 0.023817, respectively), but more than doubled to 0.053571 in the final UFS model (Table 3.2A). The final number of included covariates did not correlate with scale size, but the smallest scale (county level) did have the most covariates in the final model (n=15). Applying best-fit models from one scale to another resulted in a general pattern of higher model performance as scale decreased (Figure 3.2).

3.4.2.2. Covariates

Overall, covariates included in a final model at any given scale tended to maintain their relationship (positive or negative), estimate values, and significance in regard to the independent variable (Table 3.2B). In the UFS final model, the following covariates changed in their relationship, estimate, and/or significance when applied to the other two scales: VIlag1, asianpct, and dhipct. In the local scale final model, only the covariate dlipct changed in relationship (changing from positive to negative) and significance (changing from strongly significant to non-significant) when applied to only the UFS area. When applied to the local scale, only two covariates included in the best-fit county model changed (owpct changed from non-significant to significant; glandpct changed from negative to positive and from significant to not-significant). However, when applied to the UFS, there were numerous changes in relationships, estimates, and significances across the majority of covariates.

Within each scale's best-fit model, the overall importance (dependency) of included covariates increased with increasing scale (decreasing extent, Figure 3.3, Table 3.3). The

indicator of covariate importance, the percent change in RMSE, ranged from 0.000643 to -0.000161 in the county scale, from 0.0001344 to -0.004954 in the local scale, and from 0.003043 to -0.000467 in the UFS, a range of 0.000804, 0.00050884, and .00351, respectively. The most important covariate in the county and local scales was templag4, and templag4 was the second most important covariate in the UFS. The most important covariate in the UFS was VIlag1. Only the covariate dlipct in the local scale was notably non-important; all other negative covariate values were marginal.

3.5. DISCUSSION

3.5.1. Synthesis

Even at the largest extent, the county level was assessed at a higher spatial and temporal resolution than most WNV studies. This study's comparison provided insights into the controls and processes of WNV disease dynamics, highlighting the changes across different scales (Meentemeyer et al. 2012). These changes are most notable among the key similarities and differences between best-fit final models and the scale dependency of the covariates included within each. We found that as spatial scale increased, best-fit models decreased in explaining total variance, as defined by RMSE values. Additionally, when evaluating covariates using the LOCO method, percent differences in RMSE increased as scale increased, suggesting that as spatial scale decreases, covariate importance increases. These findings align with other studies evaluating scale-dependency of ecological processes, as well as the traditionally hypothesized ecological mechanism that "factors should be most important at scales at which they vary the most, because it will be difficult to find a statistically significant correlation when independent variables have low variance" (Wiens 1989, Cohen et al.

2016). Final model selection should always be conducted using the most robust methods, but careful assessments must be made with very large-scale models, as errors in prediction may be much greater than similar errors at smaller scales.

Despite having nine fewer covariates included in its final model, as compared to the county model, the local scale's overall RMSE was the lowest among all final models. Our initial hypothesis, claiming that as scale increased the number of covariates would also increase, was not supported by the local scale's final model. While no two scales can be explained by the exact same set of parameters, the local scale may be less heterogeneous across space than the larger county model. That being said, the UFS locations are located within the local scale, and result in more included covariates (n=10). Across all scales, only 6 (dlipct, Jantemp, MIRlag4, templag3, templag4, and totalpop) of the 23 total covariates (26.1%) assessed were included in 2 or more final models. Of these, only 2 (templag3, templag4, 8.7%) were included in all three scales' final models. Percent difference in RMSE may be the most valid estimation for each covariate's importance to a single model, but total frequency across models may provide the best indicator of its importance and robustness to WNV ecological processes.

The best-fit model from all three scales performed very well, as indicated by the ROC AUC values. However, to visually quantify the main outcome, prediction of human WNV risk, disease risk maps were created for each scale (Figure 3.4). These risk maps, displaying results (by pixel) of final models from 2005-2016, display similar overall patterns of increased disease risk, but differ in intensity among specific locations as scales decrease. For example, the local county risk map displays a focal high-risk location near the center of the study area, whereas the county and UFS risk map only

show a small cluster. As scale decreased, clusters of disease risk increased in frequency, but decreased in area, resulting in a patchy distribution, a finding that aligns with bird distribution and increasing scale in a study by Hurlbert & Jetz (2007). Despite sharing the same location and containing mostly the same covariates, the predicted relationships can have subtle, but critical difference across scales. To visualize these differences, we calculated the percent difference in human WNV illness by pixel and categorized the values across a prediction performance scale (extreme underprediction to extreme overprediction, Figure 3.5). While the extreme underprediction values never dropped below 0.00032%, some extreme overprediction values exceed 1500%. Despite being the most important human arboviral disease in North America, annual human cases are generally low. Additionally, it is estimated that at least 4 out of every 5 cases are not reported, as most humans experience little to no symptoms (Semenza and Menne 2009). If reporting of actual disease incidence were improved, extreme over- and underprediction events would likely reduce.

3.5.2. Limitations

Evaluating the role of ecological processes in disease risk comes with many challenges. Realizing issues that may arise as spatial extent increases, this study attempted to reduce error when predicting human WNV illness by including data from very high spatial and temporal resolutions. The modifiable areal unit problem, a source of statistical bias when aggregating point data by arbitrary spatial zones, can result in summaries that may change interpretations of results drastically. While many models are very well constructed and perform well, environmental stochasticity and random processes can enhance response variability, particularly at large scales (Fraterrigo et al., 2020). Our

results suggest that each scale has specific, but subtle, differences that explain the variability in WNV ecological processes. However, as the extent of our study increased (maintaining grain resolution), our models captured more unexplained variance. It is possible that at the largest scale (UFS), the sampling locations did not capture an adequate representation of that of the larger study area, resulting in higher RMSE values. However, it is expected that fine scale ecological processes may have greater influence, and therefore importance, at small spatial extents. These results provide insight into the relationships that each individual covariate, and overall model performance, has across scales. Previous studies have highlighted the importance of evaluating ecological processing across multiple scales for these reasons (Levin 1992, Condeso, T Emiko, Meentemeyer 2007, Allan et al. 2009, Meentemeyer et al. 2012).

Another potential issue that arises often in eco-epidemiology is the dilution effect (Ostfeld and Keesing 2000, Loss et al. 2009). The dilution effect states that disease risk is limited or reduced due to an increase in biodiversity. Essentially, the more potential hosts or intermediate reservoirs in a given location, the less likely an individual will acquire a given zoonotic disease. The opposite phenomenon – the amplification effect - occurs when there is a lack of biodiversity, thus increasing disease risk (Keesing et al. 2006, Johnson et al. 2015). The greater Chicago area is the third most populated area in the United States and the landscape has been severely altered by humans (swamp and grassland are now mostly pavement and built-up space). The northwest suburbs, which comprise nearly all the local and ultra-fine scale study areas, have significantly more green spaces and natural areas. It is possible that species biodiversity is greater in these

locations, as opposed to many other WNV "hot spots" throughout Chicago and playing a role in reducing the number of human WNV cases.

3.5.3. Future Directions

Future studies should continue to evaluate processes of disease ecology across multiple scales and variable landscapes (rural vs. urban, northern vs. southern latitudes, built up vs. green space, etc.). As humans continue to burn fossil fuels, maintain our exponential population growth, and encroach into new habitats, understanding and predicting the future spread of infectious diseases is of paramount importance. There are numerous studies that attempt to evaluate vector-borne disease processes, but almost all occur across very small (e.g. multiple counties or greater) scales. While helpful with understanding overall disease ecology and trends across space, finer scale processes are rarely generalizable and can lead to potential bias and invalid statistical inferences.

Along with the increased emphasis for evaluating ecological processes across scales, more robust statistical methods need to be developed to quantitatively compare processes from models fit across varying datasets. A large limitation is placed on research aiming to address effects of ecological processes across scale, which often occurs over long periods of time. Model performances are easily quantified using AIC and BIC methods, and a similarly devised metric for comparing model performance across scales would be ideal.

Despite the small differences in human WNV risk across scales, we hope risk maps like these, with the statistical rigor and methods applied, will be of great use to public health and disease control personnel. Most importantly, our ultimate goal is to allow these risk maps to be dynamic and intuitive with easy-to-interact variables that are analyses ready. Future research would greatly benefit by collaborating with environmental engineers and

programmers to facilitate in developing software that can allow for this goal to be achieved.

3.6. TABLES AND FIGURES

Table 3.1. List of covariates available for analysis by scale study area.

Covariate Information				Model Name		
Designation		Description	Notation	Ultra-fine-scale	Local	County
Abiotic	Environmental	Land Cover	Proportion of developed open space	dospct	X	X	X
			Proportion of developed low intensity	dlipct	X	X	X
			Proportion of developed medium intensity	dmipct	X	X	X
			Proportion of developed high intensity	dhipct	X	X	X
			Proportion of deciduous forests	dfpct	X	X	X
			Proportion of evergreen forests	efpct	X	X	X
			Proportion of mixed forests	mfpct	X	X	X
			Proportion of barren land	blpct	X	X	X
			Proportion of shrubs	shrubspct	X	X	X
			Proportion of grassland	glandpct	X	X	X
			Proportion of pasture	pasturepct	X	X	X
			Proportion of cultivated land	clpct	X	X	X
			Proportion of woody wetlands	wwpct	X	X	X
			Proportion of herbaceous wetlands	hwpct	X	X	X
			Proportion of total forest	ftotpct	X	X	
			Proportion of total wetlands	wtotpct	X	X	
			Proportion of open water	owpct	X	X	X
			Normalized Difference Vegetation Index	NDVI	X	X	
	Weather	Temperature	Average temperature current week	tempc	X	X	
			Average temperature of one week before	templag1	X	X	X
			Average temperature of two weeks before	templag2	X	X	X
			Average temperature of three weeks before	templag3	X	X	X
			Average temperature of four weeks before	templag4	X	X	X
			Mean January temperature	Jantemp	X	X	X
		Precipitation	Average precipitation current week	preci	X	X	
			Average precipitation of one week before	precilag1	X	X	X
			Average precipitation of two weeks before	precilag2	X	X	X
			Average precipitation of three weeks before	precilag3	X	X	X
			Average precipitation of four weeks before	precilag4	X	X	X

Table 3.1. (continued)

Covariate Information				Model Name		
Designation		**Description**	**Notation**	**Ultra-fine-scale**	**Local**	**County**
Biotic	Biological	MIR one week before	mirlag1	X	X	X
		MIR two weeks before	mirlag2	X	X	X
		MIR three weeks before	mirlag3	X	X	X
		MIR four weeks before	mirlag4	X	X	X
	Minimum Infection Rate (MIR)	Average MIR current week	MIRmean	X	X	
		Difference in weekly average MIR from 12-year average	MIRdiff	X	X	
		Vector Index current week	Vector Index	X	X	
		Vector Index one week before	VIlag1	X	X	
		Vector Index two weeks before	VIlag2	X	X	
		Vector Index three weeks before	VIlag3	X	X	
		Vector Index four weeks before	VIlag4	X	X	
	Mosquito Abundance	Light and gravid trap collection mean current week	Trap_Mean	X	X	
		Light and gravid trap collection mean one week before	Trap_Meanlag1	X	X	
		Light and gravid trap collection mean two weeks before	Trap_Meanlag2	X	X	
		Light and gravid trap collection mean three weeks before	Trap_Meanlag3	X	X	
		Light and gravid trap collection mean four weeks before	Trap_Meanlag4	X	X	
	Mosquito Biting Rates (HLC)	Mosquitoes per visit	mosquitoes per visit	X		
		Culex spp. per visit	Cx per visit	X		
	Socio-demographic	Percentage of White population	whitepct	X	X	X
		Percentage of African American population	blackpct	X	X	X
		Percentage of Asian population	asianpct	X	X	X
		Percentage of Hispanic population	hispanicpct	X	X	X
		Median household income	Income	X	X	X
		Percentage of housing constructed before WWII	hpctpreww	X	X	X
		Percentage of housing constructed post WWII (1945-1969)	hpctpostww	X	X	X
		Percentage of housing constructed from 1970-1989	hpct7089	X	X	X
		Percentage of housing constructed in 1990 or later	hpctpost90	X	X	X
	Anthropogenic	Catch basin density	CB	X	X	
		Total area of building structures	bldg_footprint_area_total	X		
		Average area of building structures	bldg_footprint_area_avg	X		
		Total perimeter of building structures	Building_Footprint_peri_total	X		
		Average perimeter of building structures	Building_Footprint_peri_avg	X		
		Total area of residential lot	Residential_lot_area_total	X		
		Average area of residential lot	Residential_lot_area_avg	X		
	Land change & manipulation	Total perimeter of residential lot	Residential_lot_peri_total	X		
		Average perimeter of residential lot	Residential_lot_peri_avg	X		
		Ratio of total building area by total lot area	total_bldg_area/total_lot_area	X		
		Ratio of average building area by average lot area	avg_bldg_area/avg_lot_area	X		
		Ratio of total building perimeter by total lot area	total_bldg_peri/total_lot_area	X		
		Ratio of average building perimeter by average lot area	avg_bldg_peri/avg_lot_area	X		
		Number of buildings	buildings	X		
		Building density per $mi.^2$	bldg_density	X		
	Human population	Number of residents per building	persons_per_bldg	X		
		Total human population	totpop	X	X	X
		Mean light pollution	lightpol	X	X	
	Activity Observations	Senior Citizen Observations per visit	Senior_obs per visit	X		
		Adults Observations per visist	Adults_obs per visit	X		
		Children Observations per visit	Child_obs per visit	X		
		Male Observations per visit	Male_obs per visit	X		
		Female Observations per visit	Female_obs per visit	X		
		Total Observations per visit	Total_obs per visit	X		
N/A	Other	Year	yr	X	X	X
		Hexagon Designation	hexid	X	X	X
		Total Covariates Evaluated		**82**	**59**	**40**

Table 3.2. Matrix of overall model performance (A) and details of each model's parameters (B). Values in bold indicate details of best-fit models for each respective scale. Remaining values are details of each scales' best-fit model applied to the other two scales.

A.			**Scale Applied To**		
			UFS	*NWMAD*	*Cook/DuPage*
Best Model Applied (# Hexagons)	*UFS* (55)	*p-value*	**<0.0001**	<0.0001	<0.0001
		df	**10**	10	10
		BIC	**546.8**	2067.5	13547.8
		ROC	**0.87**	0.89	0.85
		RMSE	**0.053571**	0.024347	0.024865
	NWMAD (1019)	*p-value*	<0.0001	**<0.0001**	<0.0001
		df	6	**6**	6
		BIC	558.704	**1987.63**	13222.1
		ROC	0.82	**0.91**	0.87
		RMSE	0.053883	**0.023817**	0.024867
	Cook/DuPage (5345)	*p-value*	<0.0001	<0.0001	**<0.0001**
		df	15	15	**15**
		BIC	632.3	2079	**13161.8**
		ROC	0.85	0.92	**0.89**
		RMSE	0.053845	0.023812	**0.024872**

Table 3.2. (continued)

B.			Scale								
		Covariate	*UFS*			*Local*			*County*		
			βn+1 relationship, estimate, & *p-value*								
Best Model	*UFS*	β_0	**-**	**13.3857**	**<0.0001**	-	16.2983	<0.0001	-	17.0640	<0.0001
		templag3	+	**0.1623**	**0.0235**	+	0.1680	<0.0001	+	0.1985	<0.0001
		templag4	+	**0.2107**	**0.0026**	+	0.2849	<0.0001	+	0.2589	<0.0001
		precilag1	-	**0.0203**	**0.0428**	-	0.0082	0.0442	-	0.0042	0.0028
		precilag2	-	**0.0142**	**0.0885**	-	0.0092	0.0194	-	0.0074	<0.0001
		precilag4	-	**0.0189**	**0.0443**	-	0.0152	0.0007	-	0.0076	<0.0001
		VIlag1*	-	**0.0037**	**0.5384**	+	0.0026	0.0020	+	0.0044	<0.0001
		asianpct	+	**0.0474**	**0.0239**	-	0.0186	0.0497	+	0.0019	0.5906
		dhipct	+	**0.0165**	**0.2297**	-	0.0113	0.1198	-	0.0074	0.0004
		mfpct	-	**4.6678**	**0.3461**	-	0.6939	0.0214	-	0.2021	<0.0001
		wwpct	-	**13.8269**	**0.9951**	-	0.1784	0.0170	-	0.0463	<0.0001
	Local	β_0	-	13.0397	<0.0001	**-**	**18.2196**	**<0.0001**	-	17.5333	<0.0001
		templag3	+	0.1447	0.0381	**+**	**0.1662**	**<0.0001**	+	0.1828	<0.0001
		templag4	+	0.1767	0.0058	**+**	**0.2445**	**<0.0001**	+	0.2294	<0.0001
		MIRlag4	+	0.0103	0.0115	**+**	**0.0090**	**<0.0001**	+	0.0045	<0.0001
		totpop	+	0.0003	0.0663	**+**	**0.0007**	**<0.0001**	+	0.0003	<0.0001
		Jantemp	+	0.0257	0.6356	**+**	**0.1262**	**<0.0001**	+	0.1288	<0.0001
		dlipct	-	0.0043	0.5512	**+**	**0.0257**	**<0.0001**	+	0.0208	<0.0001
	County	β_0	+	31.0486	0.7583	-	27.0403	0.6323	**-**	**33.4730**	**0.1189**
		Yr	-	0.0213	0.6716	+	0.0046	0.8694	**+**	**0.0082**	**0.4431**
		templag2	-	0.0505	0.4964	+	0.0664	0.1115	**+**	**0.0887**	**<0.0001**
		templag3	+	0.1660	0.0299	+	0.1333	0.0020	**+**	**0.1290**	**<0.0001**
		templag4	+	0.1680	0.0144	+	0.2070	<0.0001	**+**	**0.1840**	**<0.0001**
		Jantemp	-	0.0041	0.9451	+	0.1040	0.0031	**+**	**0.1090**	**<0.0001**
		mirlag1	+	0.0035	0.4715	+	0.0040	0.0710	**+**	**0.0035**	**<0.0001**
		mirlag2	+	0.0043	0.3715	+	0.0041	0.1368	**+**	**0.0042**	**<0.0001**
		mirlag3	+	0.0071	0.0517	+	0.0051	0.0601	**+**	**0.0044**	**<0.0001**
		mirlag4	+	0.0096	0.0309	+	0.0084	<0.0001	**+**	**0.0045**	**<0.0001**
		totpop	+	0.0001	0.5648	+	0.0005	<0.0001	**+**	**0.0002**	**<0.0001**
		owpct	-	0.0763	0.4064	-	0.0289	0.5398	**-**	**0.0613**	**0.0005**
		dlipct	-	0.0110	0.1962	+	0.0195	<0.0001	**+**	**0.0168**	**<0.0001**
		dfpct	-	0.0937	0.3387	-	0.1590	0.0452	**-**	**0.0276**	**0.0096**
		glandpct	-	0.0969	0.6389	+	0.0096	0.9184	**-**	**0.0530**	**0.0392**
		hpctpost90	+	0.0094	0.3838	-	0.0002	0.9689	**-**	**0.0067**	**0.0035**

*Vector index and its associated lags were not available for the county scale models. The next closest variable, MIR, was used as a proxy.

Table 3.3. Results of leave-one-covariate-out (LOCO) performance method per respective scale's best-fit model. The greater in magnitude of the percent change in RMSE indicates overall covariate importance (if positive) or unimportance (if negative) per each respective scale's best-fit model. Covariates removed are provided in alphabetical order.

Covariate Removed	Model									# Scales Included In
	County			*Local*			*UFS*			
	RMSE	% RMSE	Δ RMSE	RMSE	% RMSE	Δ RMSE	RMSE	% RMSE	Δ RMSE	
None	0.024872	1.000000	0.000000	0.023817	1.000000	0.000000	0.053571	1.000000	0.000000	NA
asianpct							0.053582	1.000205	0.000205	1
dfpct	0.024873	1.000040	0.000040							1
dhipct							0.053562	0.999832	-0.000168	1
dlipct	0.024876	1.000161	0.000161	0.023699	0.995046	-0.004954				2
glandpct	0.024873	1.000040	0.000040							1
hpctpost90	0.024873	1.000040	0.000040							1
Jantemp	0.024879	1.000281	0.000281	0.023831	1.000588	0.000588				2
mfpct							0.053594	1.000429	0.000429	1
MIRlag1	0.024873	1.000040	0.000040							1
MIRlag2	0.024869	0.999879	-0.000121							1
MIRlag3	0.024868	0.999839	-0.000161							1
MIRlag4	0.024869	0.999879	-0.000121	0.023802	0.999370	-0.000630				2
owpct	0.024873	1.000040	0.000040							1
precilag1							0.053546	0.999533	-0.000467	1
precilag2							0.053604	1.000616	0.000616	1
precilag4							0.053605	1.000635	0.000635	1
templag2	0.024871	0.999960	-0.000040							1
templag3	0.024877	1.000201	0.000201	0.023822	1.000210	0.000210	0.053605	1.000635	0.000635	3
templag4	0.024888	1.000643	0.000643	0.023849	1.001344	0.001344	0.053608	1.000691	0.000691	3
totpop	0.024869	0.999879	-0.000121	0.023820	1.000126	0.000126				2
VIlag1							0.053734	1.003043	0.003043	1
wwpct							0.053599	1.000523	0.000523	1
Yr	0.024873	1.000040	0.000040							1

FIGURE 3.1. The WNV model comparison study area, displaying the Chicago city limit and the 1 km hexagonal grid observational units within Cook and DuPage County. The models for the county, local, and ultra-fine scales comprise of all hexagons contained within both Cook and DuPage county (n=5345), the hexagons bounded by the Northwest Mosquito Abatement District (indicated by the orange and yellow hexagons, n=1019), and the hexagons indicated in yellow (n=55), respectively.

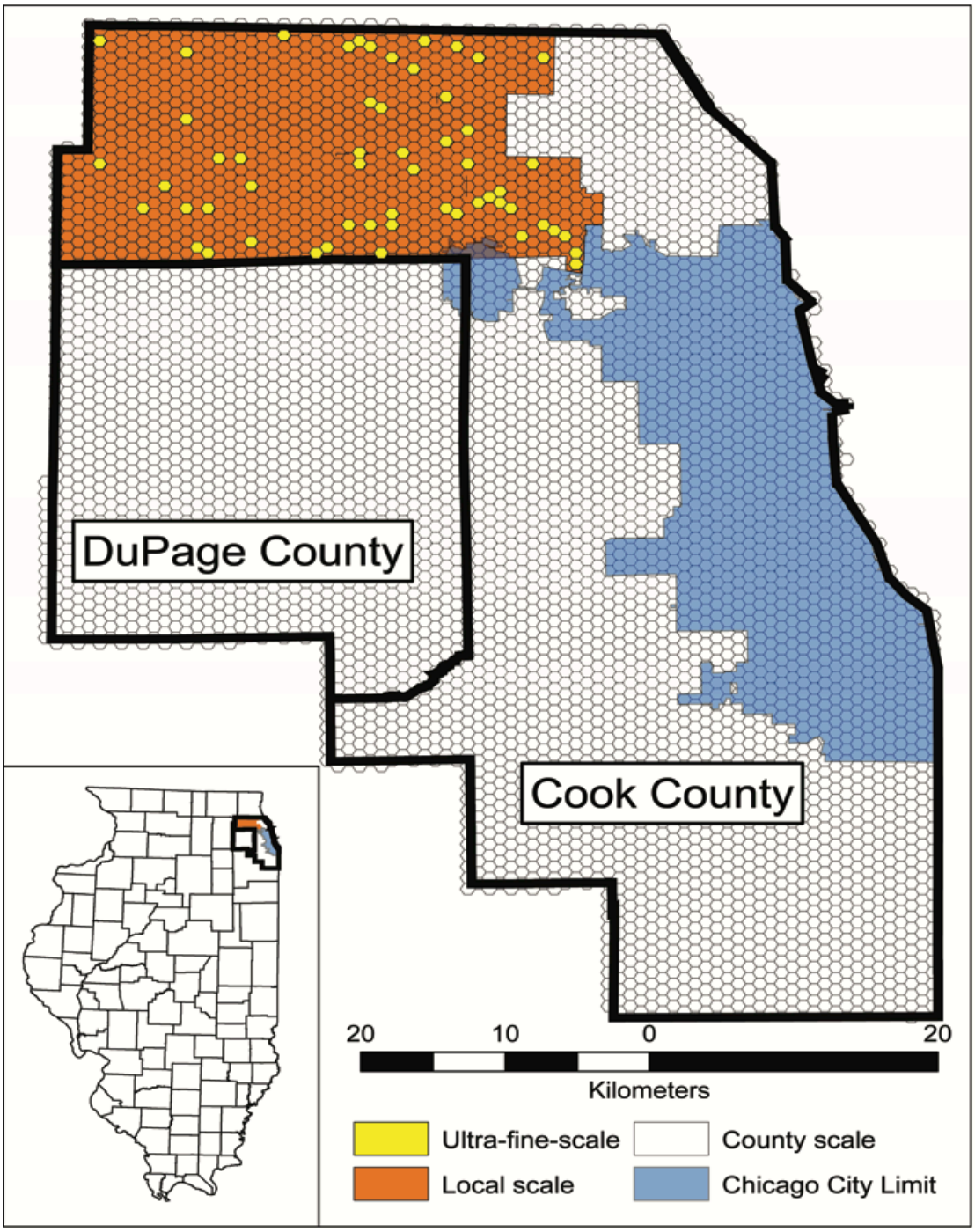


FIGURE 3.2. Overall best-fit model performances for each of the three scales. Additionally, the best-fit county (Cook/DuPage) and local (NWMAD) scale models were applied to the UFS model area (designated as larger scale model – smaller scale model, where '-' denotes 'applied to'). The best-fit county scale was also applied to the local scale model area. Models are ordered by scale size.

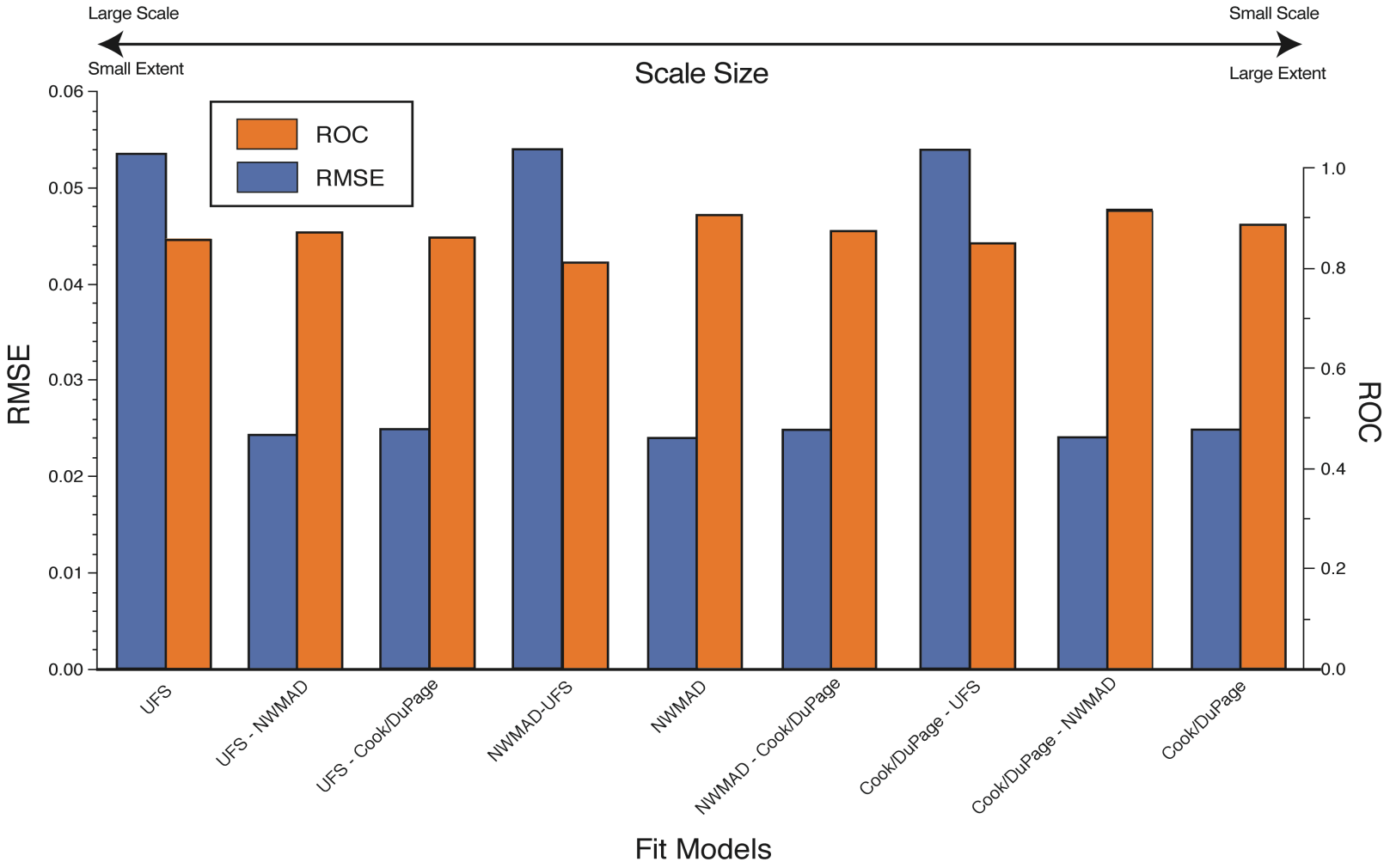


FIGURE 3.3. Evaluation of each scale's individual covariate performance indicated by the percent change in root mean square error (RMSE) of each respective best-fit model. Bars that extend further into the positive value indicated greater importance to a respective model, while bars that extend further into the negative indicated lesser importance.

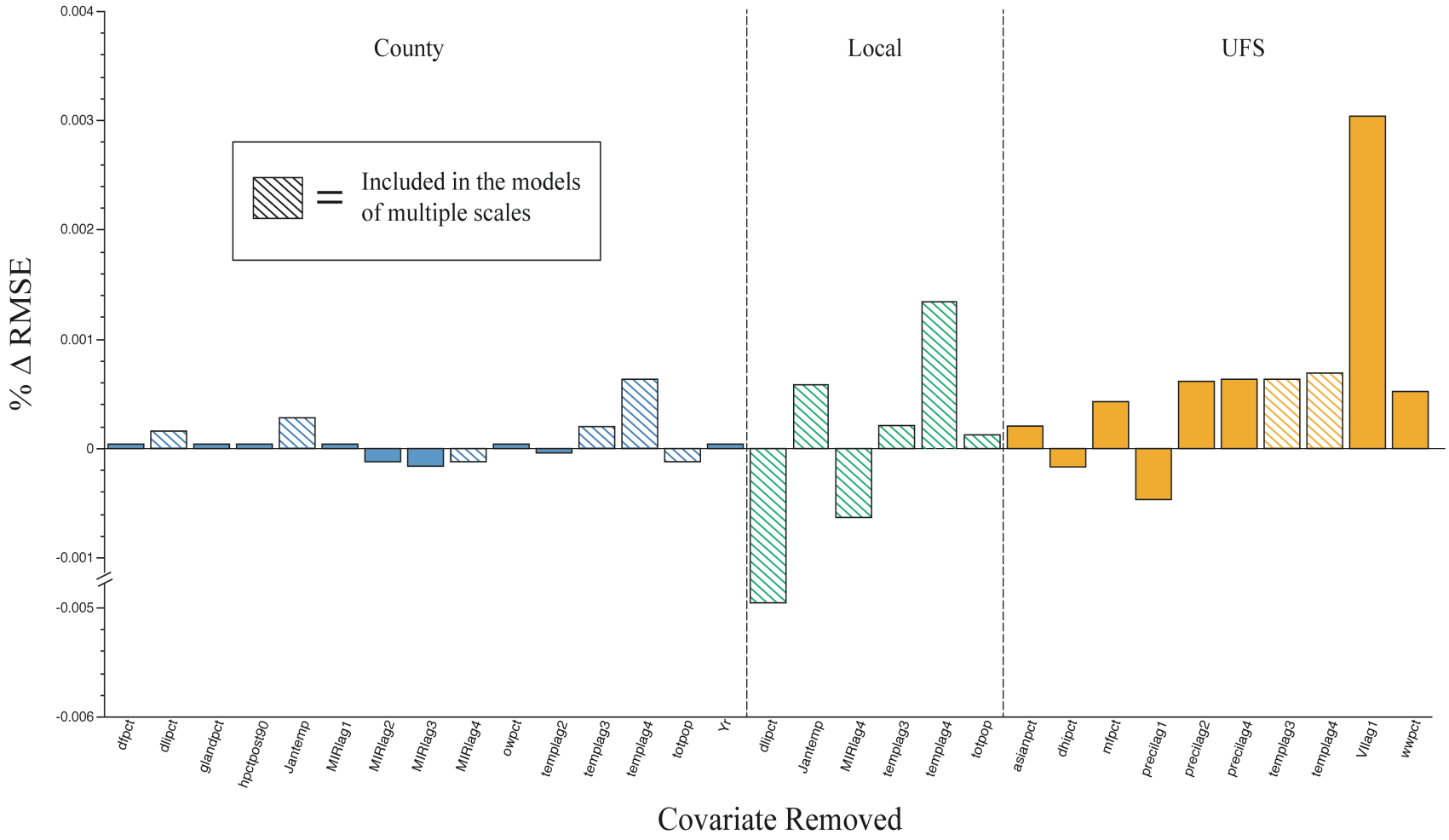


FIGURE 3.4. High-resolution human WNV risk maps for county (A), local (B), and ultra-fine (C) scales. Values correspond to total infections per 10,000 people from 2005-2016. Each pixel is populated by the best-fit models for each respective scale.

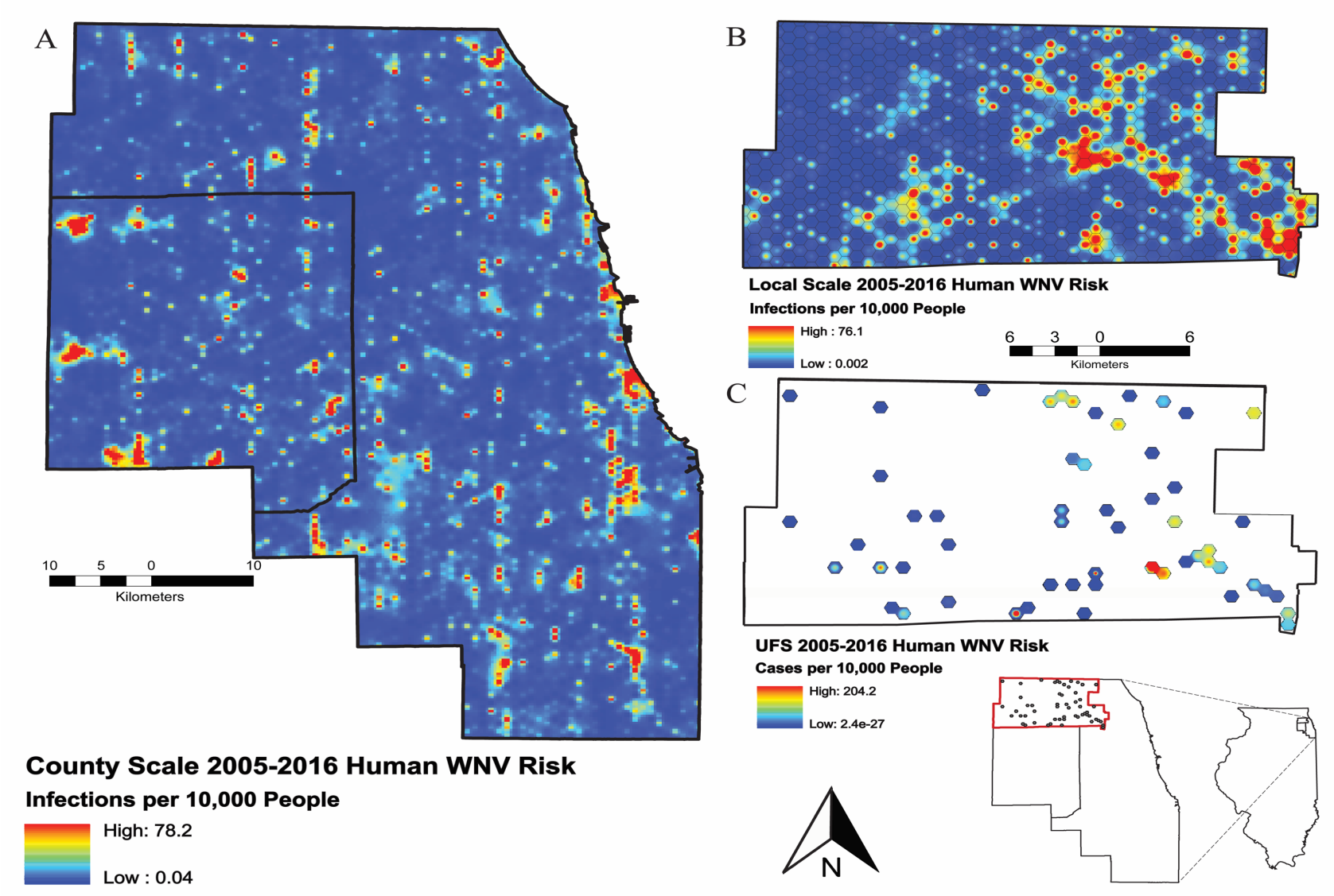
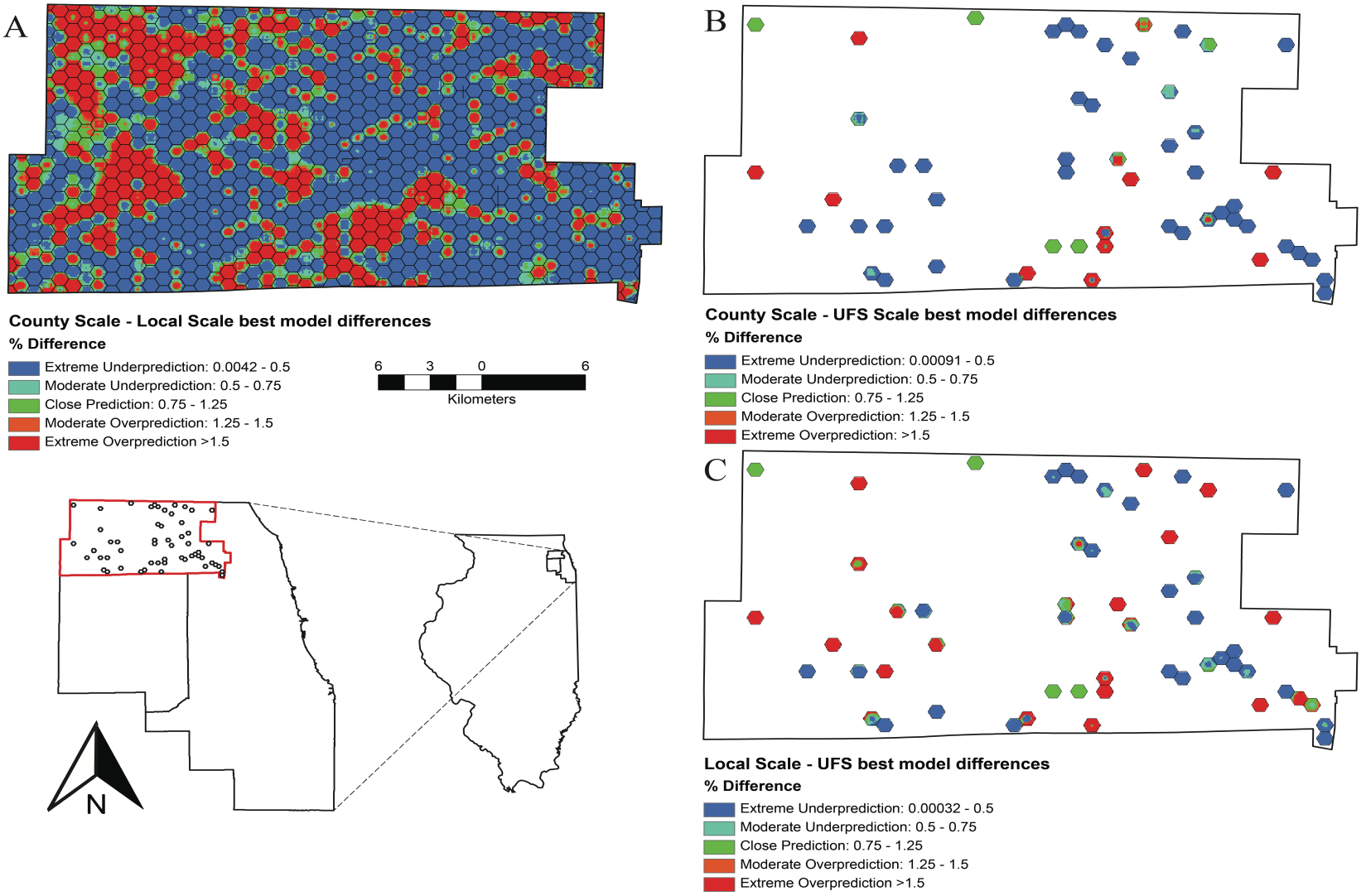


FIGURE 3.5. Model comparisons, calculated as the difference between the best-fit larger scale model and the best-fit larger-scale model, for county-local (A), county-UFS (B), and local-UFS (C).



CHAPTER 4: AN 18-YEAR RETROSPECTIVE ANALYSIS OF WEST NILE VIRUS INFECTION IN CULEX MOSQUITOES OF THE MIDWESTERN UNITED STATES

4.1. ABSTRACT

Background: West Nile virus (WNV) has been in the United States (U.S.) for over two decades, resulting in more than 50,000 human illness cases and 2,330 deaths. Now the cause for the most widespread mosquito-borne illness in humans, numerous efforts to model WNV transmission have been attempted, with widespread, and sometimes conflicting, interpretation of results. With increases in accessibility of detailed, historical data, information sharing, and collaboration across multiple partners and states, this study aims at collating and analyzing several large mosquito abundance and infection datasets from over 100 partners in the midwestern U.S. The main goal is to create a single standardized record of mosquito infection and abundance across time and space in a large region of the U.S. to effectively compare and contrast key model parameters and to generate robust forecasts under anticipated future shifts in climate.

Methods: Partners from 118 counties, representing 8 states in the Midwest U.S., provided all available records of female *Culex* spp. abundance and infection records, from 2000-2018. All datasets were organized, screened for redundancies and erroneous entries, and collated. In total, over 6.6 million mosquitoes were collected, and over 4.7 million were tested for WNV. These data were compiled to create standardized minimum infection rate (MIR) and vector index (VI) by county by month for each year. Serving as the dependent variables, MIR and VI were assessed in one of four model scenarios, with 60 explanatory variables consisting of climatic, socio-economic and demographic, and land-use/land-cover available.

Results: The vast majority of models (under any scenario) for most counties performed well, as indicated by $R^2 > 0.85$ and RMSE > 1.5. The strongest models for each county generally correlated strongly with human neuroinvasive cases. Counties with the highest resources did not generally result in higher overall mosquito infection.
Conclusions: Although most locations were predicted well, nearly all of Iowa and southern Wisconsin did not have models that correlated well with human neuroinvasive cases, suggesting knowledge gaps in environmental viral presence, particularly among rural counties. This 18-year assessment showed an annual increase in MIR by 16.4% and estimates the next outbreak year for human cases to be in 2022-2023.

4.2. INTRODUCTION

West Nile virus (WNV, Family: *Flaviviridae*) occurs in nature through an enzootic bird-mosquito vector-bird reservoir cycle (Gray and Webb 2014). When an infected mosquito, primarily from the genus *Culex*, takes a blood meal from a human, zoonotic spillover can occur (Colpitts et al. 2012). Between 75-80% of humans infected with WNV will be asymptomatic, whereas the remaining 20-25% will experience a mild fever, headache, and fatigue (CDC 2018d). However, 1 in 150, or >1% of all humans infected, will experience severe neuroinvasive illness and require urgent medical care (Carson et al. 2012).

West Nile virus arrived in the United States (U.S.) in New York City in the summer of 1999. Over the next four years, the virus rapidly spread, reaching California in 2003 (Reisen et al. 2004). The initial public health response was swift, resulting in new diagnostic testing methods, the creation of the nation's primary arthropod vector database, ArboNet (CDC 2020b), and the creation of catchy terminology like "Fight the Bite", that

accompanied other enhanced and widespread awareness campaigns focused on public education and awareness. Additionally, the establishment and implementation of routine surveillance and mosquito control has been conducted throughout numerous local, state, and federal public health and abatement agencies throughout the country (Nasci and Mutebi 2019).

Now over two decades since the virus first arrived in the U.S., WNV is now the country's most important and widespread mosquito-borne disease (Hahn et al. 2015). Despite millions of dollars in control and surveillance efforts, the virus has resulted in more than 50,000 human cases and at least 2,330 deaths (CDC 2020a). Although humans are considered incidental hosts, we can oftentimes play an integral role in propagating virus.

Combatting the virus has been extremely difficult, partly due to the complex interactions of the viral life cycle in relation to climatic factors, but also partly because there are numerous mosquito species and potential bird reservoirs that can provide suitable host-reservoir cycle combinations for the virus to maintain in the environment (Petersen and Fischer 2012). In the midwestern U.S., WNV transmission dynamics can vary by land cover type. Several *Culex* (*Cx.*) vectors, most notably *Cx. pipiens, Cx. quinquefasciatus, and Cx. restuans*, are very well suited to live in highly urban, human built environments (Andreadis et al. 2004, Ruiz et al. 2010, Gray and Webb 2014). Moreover, *Culex pipiens* form *molestus* and *molestus/pipiens* hybrids are thought to have evolved to prefer to breed in catch basins and other artificial subterranean environments (Byrne and Nichols 1999, Huang et al. 2008). However, in more rural and agricultural landscapes, particularly west of the Mississippi River, *Cx. tarsalis* is the predominant

WNV vector (Bell et al. 2005, Reisen et al. 2006). As humans continue to alter the environment, WNV seemingly thrives in nature, and is now considered endemic to North America (Ruiz et al. 2007).

Numerous efforts have attempted to model dynamics of disease, but are often focused on aspects of WNV in a particular state, cluster of counties, or specific vector species (Epstein and Defilippo 2001, Kuhn et al. 2005, Day and Shaman 2008, DeGroote et al. 2008, Wimberly et al. 2008, Ruiz et al. 2010, DeGroote and Sugumaran 2012, Kilpatrick and Pape 2013, Karki et al. 2020). Additionally, factors associated with either WNV illness in humans, prevalence in mosquitoes, or a combination of the two, oftentimes differ in studies conducted in overlapping or nearby study regions (Morin and Comrie 2013, DeGroote et al. 2014). Motivated by the absence of a standardized, cohesive approach to modeling WNV, this study evaluated available abundance and infection rates of *Culex* mosquitos provided by partners from 118 counties in 8 states of the midwestern U.S, from 2000-2018. The main objectives of this study are to: 1. Summarize and compare WNV in mosquitoes within the region, 2. Quantify and determine the importance of the main drivers associated with mosquito infection by county, 3. Evaluate any associations with surveillance efforts (or lack thereof) to mosquito infection and human illness, and 4. Generate robust models for each county based on 18 years of surveillance.

4.3. METHODS

This project did not receive nor use any personal identifiable information (PII) for any analyses. Human neuroinvasive illness data were provided by the U.S. Centers for

Disease Control (CDC) and were aggregated to cumulative annual values by county prior to receiving.

4.3.1. Study area

This study collated all available historical mosquito abundance and infection data, between 2000 and 2018, from 7 mosquito abatement districts, 7 health departments (4 county, 3 state), and 2 universities (Table 4.1A). This comprised of a total of 118 counties representing 8 states in the midwestern United States (Figure 4.1, Table 4.1B).

4.3.2. Mosquito data

Vector control, public health, and academics working in mosquito surveillance and control in the Upper Midwest were identified through collaboration with the Midwest Center of Excellence in Vector-Borne Diseases. All those who reported having data on mosquito surveillance were asked to provide any historical *Culex* spp. abundance and/or infection data. Descriptive statistics were conducted for all counties that submitted mosquito data. Several counties do not reliably record mosquito abundance information, and as a result, the total number of mosquitoes tested may sometimes exceed the reported total number of mosquitoes collected.

All submitted mosquito abundance and infection data were digitized (if paper copies were received) and processed for collation by epidemiological week for each year. Once collated, all submissions were thoroughly screened for duplicates and overall quality in reporting. Any ambiguous or extreme results (e.g. MIR >500) were removed prior to analysis.

4.3.3. Model covariates overview

This analysis evaluated sixty independent variables derived from a variety of abiotic and biotic factors, including climatic and meteorological records, mosquito infection, socio-demographic census data, and other biological conditions (described below). Each of the independent variables were calculated and aggregated by county.

4.3.3.1. Independent variable descriptions

4.3.3.1.1. Human illness

Human WNV neuroinvasive cases were provided by the U.S. CDC as cumulative values from 2000-2018 by county. For this study, we only used human illness data as choropleth county-level maps used primarily as a summary of the distribution in severe human illness.

4.3.3.1.2. Abiotic predictors

Land Cover: The 2016 U.S. Geological Survey (USGS 2016) National Land Cover Database (NLCD) provides 30 m resolution land-use/land-cover, imperviousness, and tree canopy data, presented as three classified rasters for the conterminous U.S. Where each of the 118 counties for this study overlapped a raster, the data was clipped, extracted, tabulated, and transformed into proportion (by area) by unique feature type (e.g. specific land-cover type) using the tabulate area tool in ArcGIS 10.5.1 (Environmental Systems Research Insititute 2011).

Weather: Historic monthly minimum, maximum, mean, and sum for temperature (ºC), dewpoint (ºC), and precipitation (mm) were acquired from the PRISM Climate Group (Oregon State University 2019), provided as 4-km resolution grids. As a proxy for winter temperature, monthly averages for January from 2000-2018 were incorporated. One-

month lags for temperature and mean precipitation were also included for each county. Each mean monthly value was extracted from climate grid to county using the zonal statistics as table function in ArcGIS.

4.3.3.1.3. Biotic predictors

All mosquito abundance and infection data were acquired from authorized mosquito reporting agencies, listed as either a mosquito control district, university, or health department.

Mosquito infection: A vast majority of submissions (>75%) provided minimum infection rates (MIR) by epidemiologic week. The remaining ~25% of infection data were provided as numbers of mosquitoes tested by week. In most of these instances, pool size was not provided. To provide a consistent and standardized method of calculating MIR across all counties from 2000-2018, every mosquito pool was assumed to consist of 50 mosquitoes, unless otherwise provided. Therefore, the process to calculate and aggregate MIR values by month for each county was as follows:

Step 1. If weekly MIR values already provided, then do not alter,

Step 2. If $n_{\text{mosquitoes tested}}$ is known, then: $\frac{\#\ of\ positive\ mosquito\ pools}{n_{mosquitoes\ tested}}$ x 1000,

Step 3. If $n_{\text{mosquitoes tested}}$ is not known, then: $\frac{\#\ of\ positive\ mosquito\ pools}{(50\ x\ each\ positive\ pool)}$ x 1000,

Step 4. Categorize each epidemiologic week by year to corresponding month. In instances where the epidemiologic week carried over to another month, it was designated as the week in the month that contained 4 or more days in that epidemiologic week. Once all mosquito infection data is converted to MIR and

categorized to month for each specific county*epidemiologic week*year combination, then mean monthly MIR = $\frac{\sum MIR}{n_{MIR}}$.

Mosquito abundance: Any historical female *Culex* spp. mosquitoes submitted by week by county were incorporated into this analysis as an additional measurement of WNV infection: the vector index (VI). Traditionally, the VI expresses the risk of WNV transmission of a specific vector species population by incorporating species presence, species density, and species infection (CDC 2013). For the purposes of this study, the VI was modified to incorporate all vector species within the *Culex* genus by month for each county, calculated using the following steps:

Step 1. Aggregate all female *Culex* mosquito abundance by month for each county (following methods in Step 4 of MIR calculations),

Step 2. Calculate monthly VI: $\sum_{i=Culex\ \text{spp.(pooled)}} \overline{N}_i \widehat{P}_i$,

where $\overline{N}_i$ = average density (number of mosquitoes per month) and $\hat{P}_i$= monthly MIR (proportion of mosquito pools testing positive for WNV).

Demographic: Total population, racial composition (White, African American, Hispanic, Native American, and Asian), and educational attainment (number of high school and bachelor level degrees) at the county level were extracted from the 2010 U.S. Census and converted to proportions of total. Household income, provided by the 2015 American Community Survey, was calculated as a percentage of households below and above the 2020 poverty thresholds (income based on family of 4), for each county.

4.3.4. Statistical methods

4.3.4.1. Location selection

Only counties that provided at least one positive mosquito pool between years 2000-2018 and were included in the predictive analyses. Prior to any predictive analyses, mean MIR values by county (from 2000-2018) were first compared using Tukey's HSD methods.

4.3.4.2. Model comparisons

Monthly mean mosquito infection was assessed by standard least squares regression, conducted under four model scenarios:

1. Null model, defined as:

$$Y_{a,b} = \beta_0 + \beta_1 \text{County} + \varepsilon_i,$$

2. Global model, defined as:

$$Y_a = \beta_0 + \beta_1 X_1 + \beta_{n+1} X_{n+1} + \beta_{31} X_{31} + \varepsilon_i,$$

$$Y_b = \beta_0 + \beta_1 X_1 + \beta_{n+1} X_{n+1} + \beta_{31} X_{31} + \varepsilon_i,$$

3. County-specific global model; same parameters as model scenario 2, but assessed by block (County),

4. County-specific best-fit model, defined as:

$$Y_a = \beta_0 + \beta_1 X_1 + \beta_{n+1} X_{n+1} + \varepsilon_i,$$

$$Y_a = \beta_0 + \beta_1 X_1 + \beta_{n+1} X_{n+1} + \varepsilon_i,$$

where Y_a = MIR, Y_b = VI. Model scenarios 1 and 2 were applied to all counties collectively and generated one outcome. Model scenarios 3 and 4 were applied to each individual county, defined as a block level, in each respective analysis. The same parameters from model scenario 2 were the applied to model scenario 3, regardless of

performance outcome. Model scenarios 2 and 4 were performed by forward selection process, based on lowest Akaike information criterion (AIC) value.

Each model scenario evaluated monthly mosquito infection as one of two dependent variables: MIR and VI. All 60 independent variables were available to use for both dependent variables, with the exception of MIRlag1 (1 month lag of MIR) and abundance, which was available only to models evaluating MIR, and VIlag1 (1 month lag of VI), which was available only to models evaluating VI. Regression analyses were analyzed using the Fit Model feature in JMP 14.2.0 (SAS Institute Inc. Cary, NC, USA).

Since each of the four model scenarios is assessed under varying independent variables and selection criterion (e.g. forward-selection AIC vs. no selection criterion), models may be compared primarily by root mean square error (RMSE), with f-value, R^2, and AIC statistics available as secondary evaluators. Additionally, predicted value outputs by month for model scenarios 1-3 are assessed as differences from model scenario 4.

4.3.4.3. Covariate comparisons

The overall comparison of included independent variables was best-assessed under model scenario 4, the county-specific best-fit models, as these provided the most reliable prediction values and serve as the baseline for evaluating covariate strengths. These values were assessed by county (most high-performing included covariates) and by covariate (overall greatest strength of fit across all county models) based on mean LogWorth. LogWorth is calculated as the $-\log_{10}$(p-value); the more significant the covariate is to the model, the higher the LogWorth value. LogWorth is a simple and effective method for quantifying strength of covariates in large models, especially with many strongly significant covariates generated.

4.3.4.4. Model-based WNV maps

Maps displaying cumulative mean monthly mosquito infections for MIR and VI by county were generated from each of the four model scenarios over the 18-year study period in ArcGIS (10.5.3). Mosquito infection values were interpolated via inverse distance weighted (IDW) nearest neighbor methods based on the Euclidean distances of county centroids.

4.4. RESULTS

4.4.1. Location Descriptions

Of the 118 counties that submitted mosquito abundance and/or infection data between the years 2000-2018, 52 (44.1%) reported at least one positive mosquito pool. The total number of collected female *Culex* mosquitoes submitted for this analysis was 6,610,507. An estimated total of 4,710,953 female *Culex* mosquitoes were tested across the midwestern U.S. region over the same time period (Table 4.1B).

The top 10 counties with the most collected female *Culex* mosquitoes were (in descending order): 1. Cook, IL (n=3,188,780), 2. Allen, IN (n=152,716), 3. Hamilton, IN (n=137,546), 4. Marion, IN (n=133,447), 5. Lucas, OH (n=117,632), 6. Ramsey, MN (n=82,176), 7. Hennepin, MN (n=79,008), 8. Polk, IA (n=40,562), 9. Macon, IL (n=37,753), and 10. St. Louis, MO (n=29,332). The top 10 counties with the most female *Culex* tested were (in descending order): 1. St. Louis, MO (n=1,291,254), 2. Cook, IL (n=1,069,768), 3. Macon, IL (n=306,958), 4. Allen, IN (n=146,460), 5. Hamilton, IN (n=148,831), 6. Saginaw, MI (n=107,168), 7. Marion, IN (n=81,710), 8. Tuscola, MI (n=61,129), 9. Bay, MI (n=52,073), and 10. Milwaukee, WI (n=51,846).

Prior to model fitting and predictive analyses, non-transformed cumulative monthly means from 2000-2018, by county, were statistically compared using Tukey's HSD. Analysis shows a broad categorization of eight levels, ranging from 9 counties (7.6%) that are uniquely different from each other location to 66 counties (56%) with no statistically significant difference among each other or any other county (Table 4.2).

4.4.2. Model Comparison and WNV Predictions

County-specific model-fit statistics were compared under scenarios 3 (county-specific global, Table 4.3A) and 4 (county-specific best-fit, Table 4.3B). The top three best-fit counties (in descending order) were Dubuque, IA, Huron, MI, and Monona, IA for model scenario 3 MIR and Pottawattamie, IA, Dane, WI, and Woodbury, IA for model scenario 3 VI. The top three best-fit counties (in descending order) were Hamilton, IN, Dubuque, IA, and Marion, IN for model scenario 4 MIR and Dubuque, IA, Carroll, IA, and Pottawattamie, IA, for model scenario 4 VI.

Prediction estimates of MIR and VI for each county, by model scenario, are provided as means and cumulative values (Table 4.4). The prediction estimates for mean monthly MIR were consistently highest across each of the four model scenarios in the counties of Lucas, OH, Ramsey, MN, and Anoka, MN. The prediction estimates for mean monthly VI were consistently highest in Lucas, OH, Cook, IL, and Ramsey, MN counties. These consistencies in the highest infected counties are evident geographically, depicted as dark shades of red in regional maps (Figure 4.2). Further assessments of model comparisons can be evaluating in Appendix C, Table 1, comparing differences in estimates of each of model scenarios 1-3 from model scenario 4 (designated as the baseline for comparison).

The cumulative (2000-2018) monthly mean MIR and VI values, predicted from each of the four model scenarios, display similar trends in "hot" and "cold" spots of MIR, or regions of high and low estimates values of mosquito infection in relation to one another (Figure 4.3). The "hot" spot regions are centered in the Minneapolis/St. Paul, Chicago, and Toledo/Detroit metropolitan areas, with the remainder of the region largely surrounded by "cold" spots of estimated low mosquito infection.

4.4.3. Covariate Performance

Using model scenario 4 (county-specific best-fit model) as the designated baseline, overall covariate performance was standardized and assessed by county for MIR and VI outcomes by LogWorth transformation (Table 4.5). Overall, the three covariates with the highest mean LogWorth were 1-month MIR lag (mirlag1), month, and mean monthly precipitation (meanppt). The top three counties with the highest mean LogWorth, for all included covariates in their final models, were Hamilton, IN, Marion, IN, and Milwaukee, WI and Lucas, OH, Hamilton, IN, and Milwaukee, WI for outcomes MIR and VI, respectively.

4.5. DISCUSSION

This study provides an ambitious analysis of WNV, reported as mean monthly MIR and VI values by county. In addition to working across multiple agencies from 8 states, each with their own unique methods of collecting and reporting data, the processing, quality checking, and standardizing of multiple very large historical data sets proved to be a massive effort, and none were without limitations. Fortunately, results from this study have provided clarity in regional and large-scale modeling of WNV, reinforcing key characteristics and features that either propagate or mitigate the virus in the environment.

Mosquito control and surveillance efforts vary in scale and size across the Midwest U.S. Counties that have dedicated abatement districts generally have the most resources that result in the richest mosquito datasets available. However, increased control efforts do not necessarily translate into higher infection rates. For example, Cook, IL, and Allen, Hamilton, and Marion, IN counties provided the highest *Culex* abundance data, but the highest cumulative mean MIR among the four was a modest 4.38 (Cook, IL, Tables 4.1B & 4.2). Nonetheless, mosquito control of any kind is linked to reductions in entomological indicators of WNV transmission (Nasci and Mutebi 2019), making the case that routine surveillance activities should be prioritized, especially in rural counties where our understanding in the ecology, transmission, and overall risk of the virus is severely lacking.

In an effort to address the lacking data in rural counties, we extrapolated the results from each of the four model scenarios to create continuous fluid maps predicting both cumulative mean monthly MIR and VI values for the entire 8-state midwestern U.S. region. Despite inputs from different models, each map prediction displays similar trends in "hot" and "cold" spots of MIR, or regions of high and low estimates values of mosquito infection, supporting modeling techniques applied to this disease system.

These values are simple IDW predictions, but are derived from complex models powered by 18-years of surveillance and a large geographic expanse. Undoubtedly, these prediction maps can be improved, particularly among counties with little or no surveillance records. As reported in this study, these locations provided prediction values that may experience the most drastic changes if future surveillance is conducted. That being said, these results differ from reported human neuroinvasive illness, particularly in

Iowa and Wisconsin. Cumulative human neuroinvasive data (CDC 2020b) from the same time period as this study's analysis shows that nearly half of Iowa's counties and Wisconsin's two most populous counties, Dane and Milwaukee, are in the upper 50% of the Midwest's human cases (by both total number and incidence rate, Figure 4.4). However, maps generated from this study denote the entire state of Iowa as a state with low mosquito infection prevalence and only account for northern Wisconsin as an area with moderate infection prevalence. This suggests that surveillance efforts and/or data reported are lacking or otherwise not representative of the true infection rate in mosquitoes in these counties

While the true number of infected mosquitoes collected is likely different than what was reported for this study, the numbers are believed to be the most accurate representation of historic mosquito infection in the midwestern U.S., considering the amount of resources that are used to acquire such information. Unfortunately, the same cannot be said for mosquito abundance numbers. The agencies reporting this data are reputable and reliable, but standardized collections and accurate identification of mosquitoes across agencies are generally not prioritized, nor consistent in methodology. The main goal for mosquito control agencies is to gather an accurate representative sample of infected mosquitoes in nature, as a function of limited surveillance efforts. The sheer numbers of mosquitoes collected often overwhelm agencies and are generally not identified (limited to genus level, if at all) or counted and are often discarded. While there is no fault or harm in these processes, the true number of mosquitoes that have been collected is undoubtedly much higher than what has been reported here and should be made known. Nonetheless, the

number of collected female *Culex* mosquitoes reported over the 18-year period, across 118 counties, exceeds 6.6 million – a remarkable effort.

Weighted by LogWorth, a $\log_{10}$-transformation of p-values for covariates included in final models, this study found that the 1 month lag of MIR, month, and mean monthly precipitation were the top three most meaningful factors in explaining mosquito infection. With the exception of percent high school and bachelor level degrees by county, no other socio-economic, demographic, or land-use/land-cover variable was included in any final model. This is likely the result of an ecological process that decreases explanatory power for fine-scale processes when evaluating outcomes at larger-scales. For example, in Karki et al. (2020) and Uelmen et al. (2020a) both studies evaluated human WNV illness at 1-km or smaller hexagonal units at weekly intervals in the Chicago, IL region, and found several indicators of biotic and human-derived (socio-economic and demographic) factors increasingly meaningful at these scales. However, in another study of the same geographic region, Uelmen et al. (2020) also found that as spatial extent decreased, fewer biotic and human-derived variables were included in final models.

By evaluating LogWorth as a cumulative mean by county, this study was able to quantify how well overall best-fit models captured variance. As a combination of the number of included significant covariates and overall explanatory power per covariate, the counties with the highest cumulative mean LogWorth values are exemplary locations in mosquito control and reporting. With the exception of Lucas County (OH), the highest cumulative mean LogWorth counties were not the counties with the highest overall mean MIR, suggesting that these counties captured and reported a more representative sampling of WNV in the environment than other counties.

4.5.1. Climate Change Implications

While the retrospective analyses of WNV in the 8-state midwestern U.S. region has been insightful, the main motivation for these analyses was to derive county- and region-specific best-fit models that can forecast WNV in the environment. Based on the 18-year dataset provided from the 118 counties in this study, mean WNV is increasing in mosquito pools by an estimated 14.2% each year (Figure 4.5). Outbreaks of human illness years coincide with years of highest annual mean MIR, occurring in 2005-2006, 2012-2013, and 2018. The gaps in between these three periods are being reduced by about 1 year, and if the pattern maintains its course, the next outbreak will occur in 2022-2023.

Increases in human surveillance can account for a minor aspect in the explanation of the increasing pattern of WNV in mosquitoes. The likely culprit, however, are changes in climate. In this analysis, the majority of models are heavily driven by abiotic forces (namely climatic drivers) in comparison to land-use/land-cover and other biotic forces (e.g., socio-economic and demographic factors). Understanding the connection to, and evaluation of, impending consequences of changing climatic effects on vector-borne diseases must be evaluated (Morin and Comrie 2013). Next steps for this study are to apply the models derived in this study to future projected climate changes over the next several decades using IPCC climate simulations specific to the upper midwestern U.S. Forecasting models will also include projected changes as a result of anthropogenic disturbances (e.g., population growths and changes to land-cover/land-use). The purposes for evaluating regional climate change effects on WNV prevalence is to provide guidance to public health officials, mosquito control personnel, and policy-makers so that proper

changes can be made to increase preparedness and mitigate human infections in future environments that are likely to result in increases in WNV in mosquitoes and birds.

Lastly, the ultimate goal of this study is to create accurate and reliable forecasting models that can provide public health and mosquito control personnel lead time to mitigate illness in humans. Although 1-month lags of MIR were statistically significant in predicting future mosquito infection, to translate this into an effective public health response would require rapid disease testing and data sharing. The 1-month window to effectively do this is incremently shortened with variable human incubation periods, disease severity, and the healthcare resources and knowledge to correctly and rapidly suspect arboviral infection as the cause of illness.

This study was made possible by the willingness to share and collaborate data across multiple agencies, and is an excellent example displaying how effective data sharing can lead to new insights and trends. We sincerely hope this will inspire and motivate increases in collaboration and data sharing across disciplines in the future.

4.6. TABLES AND FIGURES

Table 4.1. Sources of mosquito data arranged by agency type and state of origin (A). Values in parenthesis indicate available time period for each data type. Breakdown of provided data by county for each state (B). Data not provided labeled as N.P.

A.

Name	Agency Type	State	Mosquito Data		
			Abundance	WNV Infection	
				Negative Pools	Positive Pools
Desplaines Valley	Mosquito Abatement District	IL	✓ (2002-17)	✓ (2002-03, 2005-16)	✓
DuPage	Mosquito Abatement District	IL		✓ (2005-19)	✓
North Shore	Mosquito Abatement District	IL	✓ (2002-16)	✓ (2005-16)	✓
Macon	Mosquito Abatement District	IL		✓ (2002-18)	✓
Northwest	Mosquito Abatement District	IL	✓ (1992-16)	✓ (2001-03, 2005-16)	✓
Illinois Department of Health	State Health Department	IL		✓ (2019)	✓
Allen	County Health Department	IN		✓ (2003-19)	✓
Hamilton	County Health Department	IN	✓ (2015-19)	✓ (2014-19)	✓
Marion	County Health Department	IN	✓ (1989-19)	✓ (2014-19)	✓
Iowa State	University	IA	✓ (1967-19)	✓ (2002-19)	✓
Multiple County Health Departments	State Health Department	MI		✓ (2006-19)	✓
Metropolitan Mosquito Control	Mosquito Abatement District	MN	✓	✓ (2005-2019)	✓
Minnesota Department of Health	State Health Department	MN	Partial	Partial (2003-04)	Partial
St. Louis	County Health Department	MO	✓	✓ (2005-18)	✓
Toledo Area Sanitary District	Mosquito Abatement District	OH	✓	✓ (2014-19)	✓
University of Wisconsin	University	WI	✓	✓ (2009-19)	✓

Table 4.1. (continued)

B.

County	State	N	# Collected Female *Culex* Mosquitoes	# Female *Culex* Mosquitoes Tested
Allamakee	IA	13	*N.P.*	274
Appanoose	IA	22	*N.P.*	95
Audubon	IA	15	*N.P.*	571
Black Hawk	IA	1402	4426	3522
Boone	IA	130	2661	45
Butler	IA	22	*N.P.*	162
Carroll	IA	574	1577	3047
Cerro Gordo	IA	329	502	834
Cherokee	IA	7	*N.P.*	234
Clarke	IA	2	*N.P.*	35
Clayton	IA	3	*N.P.*	33
Dallas	IA	36	151	174
Davis	IA	34	*N.P.*	249
Decatur	IA	13	*N.P.*	274
Des Moines	IA	3	*N.P.*	22
Dickinson	IA	74	252	105
Dubuque	IA	921	2245	990
Emmet	IA	11	*N.P.*	267
Fremont	IA	78	*N.P.*	635
Guthrie	IA	3	*N.P.*	13
Hancock	IA	107	*N.P.*	4503
Hardin	IA	1	*N.P.*	5
Henry	IA	4	2	2
Jefferson	IA	55	454	453
Johnson	IA	68	*N.P.*	642
Kossuth	IA	4	*N.P.*	31
Lee	IA	82	277	414
Linn	IA	1290	3537	1699
Louisa	IA	7	*N.P.*	129
Lucas	IA	150	*N.P.*	2001
Lyon	IA	15	6	13
Marion	IA	20	40	40
Marshall	IA	133	251	738
Monona	IA	434	217	4392
Montgomery	IA	6	*N.P.*	77
Muscatine	IA	168	835	22
O'Brien	IA	409	341	5723
Page	IA	41	*N.P.*	397
Polk	IA	13631	40562	41408
Pottawattamie	IA	1741	3080	24141
Ringgold	IA	160	200	905
Sac	IA	58	*N.P.*	589
Scott	IA	2406	7191	14904
Shelby	IA	45	625	611
Sioux	IA	826	4701	3201
Story	IA	6947	26031	50910
Taylor	IA	4	*N.P.*	79
Union	IA	2	*N.P.*	9
Van Buren	IA	89	*N.P.*	2019
Wapello	IA	391	1142	960
Warren	IA	14	*N.P.*	614
Washington	IA	71	*N.P.*	1016
Wayne	IA	11	*N.P.*	54
Webster	IA	3	*N.P.*	15
Winneshiek	IA	3	*N.P.*	21
Woodbury	IA	4063	17720	23835
Worth	IA	8	*N.P.*	132
Champaign	IL	158	*N.P.*	1509
Cook	IL	41898	3188780	1069768
DuPage	IL	814	*N.P.*	28504
Macon	IL	15043	37753	306958
Allen	IN	4739	152761	146460
Hamilton	IN	4035	137546	148831
Marion	IN	2644	133447	81710
Arenac	MI	296	*N.P.*	1158
Barry	MI	62	*N.P.*	166
Bay	MI	2889	*N.P.*	52073
Calhoun	MI	6	*N.P.*	12
Cass	MI	6	*N.P.*	18
Genesee	MI	581	*N.P.*	3574
Gladwin	MI	72	*N.P.*	389
Huron	MI	114	*N.P.*	2269
Ingham	MI	564	*N.P.*	5309
Iosco	MI	48	*N.P.*	209
Isabella	MI	129	*N.P.*	1495
Kalamazoo	MI	16	*N.P.*	23
Kent County	MI	83	*N.P.*	3597
Lapeer	MI	1	*N.P.*	7
Livingston	MI	78	*N.P.*	275
Macomb	MI	7	*N.P.*	110
Midland	MI	148	*N.P.*	3691
Oakland	MI	424	*N.P.*	4013
Saginaw	MI	19084	*N.P.*	107168
Tuscola	MI	2823	*N.P.*	61129
Wayne	MI	78	*N.P.*	1407
Anoka	MN	4996	27338	13811
Blue Earth	MN	21	1151	49
Calhoun	MN	1	3	*N.P.*
Carlton	MN	47	948	*N.P.*
Carver	MN	4167	22879	10337
Chisago	MN	179	621	153
Clearwater	MN	9	10	*N.P.*
Dakota	MN	9194	49601	24039
Douglas	MN	20	774	*N.P.*
Freeborn	MN	9	928	*N.P.*
Hennepin	MN	12032	79008	40590
Houston	MN	36	86	*N.P.*
Isanti	MN	7	8	*N.P.*
Le Sueur	MN	546	3275	1392
Mille Lacs	MN	62	1202	*N.P.*
Morrison	MN	37	95	*N.P.*
Murray	MN	111	8026	*N.P.*
Ramsey	MN	10101	82176	58310
Renville	MN	93	9676	*N.P.*
Rice	MN	404	3167	141
Scott	MN	5974	50094	19928
Sherburne	MN	86	330	89
Sibley	MN	113	493	272
St. Louis	MN	6	25	*N.P.*
Stearns	MN	8	298	*N.P.*
Steele County	MN	79	325	*N.P.*
Washington	MN	3212	11256	4196
Winona	MN	19	158	*N.P.*
Wright	MN	696	2853	573
St. Louis	MO	29332	*N.P.*	1291254
Lucas	OH	3226	117632	24650
Dane	WI	1679	11378	8819
Milwaukee	WI	2744	47878	51846
Ozaukee	WI	12	28	28

Table 4.2. Comparison of cumulative monthly mean MIR by county between 2000-2018, assessed by Tukey's HSD methods. Values not connected by the same letter are significantly different.

County	Mean Monthly MIR (2000-2018)	Tukey's HSD	Interpretation
Boone*	66.67	A	Uniquely different from all other counties
Ramsey	12.22	B	
Genesee	0.06	G	
Chisago	0	G	
Rice	0	G	
Arenac	0	G	
Le Sueur	0	G	
Livingston	0	G	
Wright	0	G	
Anoka	11.81	BH	Uniquely different from many other counties
Macon	1.67	FG	
Pottawattamie	1.34	FG	
Story	1.34	FG	
Saginaw	1.29	FG	
Black Hawk	1.22	FG	
Woodbury	1.14	FG	
Gladwin	0.47	FG	
Dane	0.28	FG	
Dubuque	0.15	FG	
Sherburne	0	FG	
Linn	0	FG	
Lucas	11.84	BCH	Differs from about 2/3 of counties
St. Louis	2.39	EFG	
Polk	1.9	EFG	
Bay	1.63	EFG	
Tuscola	1.37	EFG	
Iosco	0.32	EFG	
Hennepin	10.32	BCHI	Differs from more than half of all other counties
Allen	3.1	DEFG	
Marion	2.26	DEFG	
Milwaukee	1.08	DEFG	
Cerro Gordo	0.06	DEFG	
Washington	9.42	BCDHI	Different from less than half of all other counties
Isabella	1.81	DEFGI	
Hamilton	1.69	DEFGI	
Carroll	0.38	DEFGI	
Sibley	0.18	DEFGI	
Wapello	0.01	DEFGI	
Johnson	0	DEFGI	
Scott	7.47	BCDEHI	Differs from about 1/3 of counties
Cook	4.38	CDEFGI	
Dakota	7.63	BCDEFHI	Differs from only a few counties
Allamakee	0	CDEFGHI	
Lee	0	CDEFGHI	

County	Mean Monthly MIR (2000-2018)	Tukey's HSD	Interpretation
Renville	15.29	BCDEFGHI	Does not differ from any other county
Kent County	13.72	BCDEFGHI	
Webster	10.75	BCDEFGHI	
Macomb	7.79	BCDEFGHI	
Wayne	7.33	BCDEFGHI	
Ingham	6.35	BCDEFGHI	
Carver	5.77	BCDEFGHI	
Huron	5.50	BCDEFGHI	
Sioux	4.52	BCDEFGHI	
DuPage	4.20	BCDEFGHI	
Douglas	4.15	BCDEFGHI	
Midland	3.34	BCDEFGHI	
Monona	3.15	BCDEFGHI	
Henry	2.86	BCDEFGHI	
Mille Lacs	2.85	BCDEFGHI	
Champaign	2.51	BCDEFGHI	
Kossuth	2.43	BCDEFGHI	
Oakland	2.17	BCDEFGHI	
Fremont	1.55	BCDEFGHI	
Cass	1.48	BCDEFGHI	
Emmet	1.42	BCDEFGHI	
Dickinson	1.35	BCDEFGHI	
Sac	1.22	BCDEFGHI	
Lyon	1.13	BCDEFGHI	
Van Buren	0.77	BCDEFGHI	
Stearns	0.74	BCDEFGHI	
Hardin	0.66	BCDEFGHI	
Appanoose	0.58	BCDEFGHI	
Hancock	0.43	BCDEFGHI	
Marshall	0.38	BCDEFGHI	
Shelby	0.29	BCDEFGHI	
O'Brien	0.23	BCDEFGHI	
Dallas	0.22	BCDEFGHI	
Butler	0.00	BCDEFGHI	
Calhoun	0.00	BCDEFGHI	
Cherokee	0.00	BCDEFGHI	
Clarke	0.00	BCDEFGHI	
Clayton	0.00	BCDEFGHI	
Clearwater	0.00	BCDEFGHI	
Montgomery	0.00	BCDEFGHI	
Morrison	0.00	BCDEFGHI	
Muscatine	0.00	BCDEFGHI	
Carlton	0.00	BCDEFGHI	
Davis	0.00	BCDEFGHI	
Ozaukee	0.00	BCDEFGHI	
Page	0.00	BCDEFGHI	
Decatur	0.00	BCDEFGHI	
Des Moines	0.00	BCDEFGHI	
Freeborn	0.00	BCDEFGHI	
Guthrie	0.00	BCDEFGHI	
Ringgold	0.00	BCDEFGHI	
Houston	0.00	BCDEFGHI	
Isanti	0.00	BCDEFGHI	
Jefferson	0.00	BCDEFGHI	
Kalamazoo	0.00	BCDEFGHI	
Lapeer	0.00	BCDEFGHI	
Taylor	0.00	BCDEFGHI	
Union	0.00	BCDEFGHI	
Warren	0.00	BCDEFGHI	
Louisa	0.00	BCDEFGHI	
Audubon	0.00	BCDEFGHI	
Barry	0.00	BCDEFGHI	
Winneshiek	0.00	BCDEFGHI	
Winona	0.00	BCDEFGHI	
Blue Earth	0.00	BCDEFGHI	
Worth	0.00	BCDEFGHI	

*Based on 1 positive mosquito pool. Results are not reliable.

Table 4.3. Model fit statistics, predicting the outcome of cumulative monthly MIR and VI between 2000-2018, by county-specific global (A) and best-fit (B) modeling methods.

A. County	MIR						VI					
	p-value	*df*	*f-value*	*R^2 adjusted*	*AIC*	*RMSE*	*p-value*	*df*	*f-value*	*R^2 adjusted*	*AIC*	*RMSE*
Overall	<0.0001	33	1389.47	0.18	1530133.00	10.37	<0.0001	76	905.96	0.60	475689.50	45.33
Allen, IN	<0.0001	23	423.26	0.72	1498.10	1.72	<0.0001	23	654.12	0.80	23619.95	5.34
Anoka, MN	<0.0001	23	238.50	0.58	13577.81	1.30	<0.0001	23	44.27	0.60	2693.71	1.74
Bay, MI	<0.0001	21	398.21	0.76	8829.15	1.25	No Abundance Data to Assess					
Carroll, IA	<0.0001	2	160.45	0.22	1086.91	1.09	No Abundance Data to Assess					
Carver, MN	<0.0001	23	504.71	0.78	10127.82	1.12	<0.0001	23	64.16	0.71	1971.61	1.22
Champaign, IL	<0.0001	7	116.03	0.85	361.99	0.83	No Abundance Data to Assess					
Cook, IL	<0.0001	45	700.10	0.18	1111995.00	12.06	<0.0001	23	1029.78	0.60	179053.60	65.24
Dakota, MN	<0.0001	23	276.28	0.46	21734.39	1.07	<0.0001	23	205.40	0.78	6442.62	2.79
Dane, WI	<0.0001	23	51.00	0.65	932.01	0.50	<0.0001	23	65.74	0.73	-861.40	0.11
Dubuque, IA	<0.0001	11	1.55^13	1.00	-3020.66	6.118^-8	No Abundance Data to Assess					
DuPage, IL	<0.0001	8	8687.35	0.99	490.57	0.45	No Abundance Data to Assess					
Hamilton, IN	<0.0001	19	3920.79	0.98	-73.30	0.23	Unable To Assess					
Hennepin, MN	<0.0001	23	742.80	0.65	26474.42	0.99	<0.0001	23	395.32	0.82	10129.62	3.02
Huron, MI	<0.0001	1	649.22	0.92	-54.74	0.15	No Abundance Data to Assess					
Ingham, MI	<0.0001	21	13.44	0.36	3869.16	14.56	No Abundance Data to Assess					
Lucas, OH	<0.0001	21	310.66	0.75	5017.62	0.75	<0.0001	10	24331.24	1.00	11586.03	3.24
Macon, IL	Unable To Assess						No Abundance Data to Assess					
Marion, IN	<0.0001	24	23.57	0.85	364.42	1.33	<0.0001	24	631.12	0.88	8281.17	1.63
Midland, MI	<0.0001	4	583.58	0.95	247.95	0.69	No Abundance Data to Assess					
Milwaukee, WI	<0.0001	20	761.11	0.95	304.62	0.28	Unable To Assess					
Monona, IA	<0.0001	3	653.36	0.99	-1.26	0.20	Unable To Assess					
Oakland, MI	<0.0001	13	620.66	0.96	272.46	0.35	No Abundance Data to Assess					
Polk, IA	<0.0001	23	413.62	0.63	14767.30	0.91	<0.0001	23	316.13	0.68	7533.56	0.71
Pottawattamie, IA	<0.0001	11	72.57	0.87	286.03	0.74	<0.0001	18	37842.36	1.00	-4940.03	0.00
Ramsey, MN	<0.0001	23	1555.40	0.83	18068.38	0.84	<0.0001	23	142.82	0.59	16004.91	7.70
Saginaw, MI	<0.0001	21	501.89	0.38	76496.99	2.29	Unable To Assess					
Scott, MN	<0.0001	24	1067.31	0.81	20820.28	1.35	<0.0001	24	145.62	0.68	9568.39	4.67
St. Louis, MO	Unable To Assess						Unable To Assess					
Story, IA	<0.0001	23	412.25	0.79	6588.37	0.87	<0.0001	23	187.62	0.62	8266.36	1.16
Tuscola, MI	<0.0001	21	374.40	0.76	9701.42	1.77	No Abundance Data to Assess					
Washington, MN	<0.0001	23	160.15	0.58	8579.35	1.24	<0.0001	24	133.87	0.91	1290.98	1.62
Wayne, MI	0.0017	12	3.46	0.37	383.05	8.12	No Abundance Data to Assess					
Woodbury, IA	<0.0001	23	196.68	0.77	3011.41	0.75	<0.0001	23	168.39	0.82	393.98	0.30

Table 4.3. (continued)

B.

County	MIR						VI					
	p-value	*df*	*f-value*	R^2 *adjusted*	*AIC*	*RMSE*	*p-value*	*df*	*f-value*	R^2 *adjusted*	*AIC*	*RMSE*
Allen, IN	<0.0001	30	461.10	0.79	13980.65	1.51	<0.0001	29	713.40	0.84	22627.69	4.68
Anoka, MN	<0.0001	28	305.60	0.70	11206.47	1.14	<0.0001	26	52.46	0.66	2570.89	1.58
Bay, MI	<0.0001	27	444.94	0.82	8069.32	1.08	No Abundance Data to Assess					
Carroll, IA	<0.0001	2	51.09	0.22	1086.91	1.09	<0.0001	6	532.17	0.88	-2685.19	0.01
Carver, MN	<0.0001	28	432.27	0.79	9622.36	1.07	<0.0001	24	69.40	0.73	1919.77	1.16
Champaign, IL	<0.0001	5	163.80	0.85	358.25	0.82	No Abundance Data to Assess					
Cook, IL	<0.0001	28	2312.65	0.80	52654.56	1.27	<0.0001	30	1028.68	0.66	176380.60	60.04
Dakota, MN	<0.0001	26	253.10	0.52	18048.44	1.05	<0.0001	27	208.04	0.81	6265.98	2.60
Dane, WI	<0.0001	25	381.96	0.90	420.91	0.29	<0.0001	25	87.29	0.79	-1016.99	0.09
Dubuque, IA	<0.0001	9	92.65	0.80	-133.69	0.17	<0.0001	15	188297.40	1.00	-1125.89	0.00
DuPage, IL	<0.0001	9	7748.39	0.99	490.53	0.45	No Abundance Data to Assess					
Hamilton, IN	<0.0001	17	84547.42	1.00	-5125.62	0.06	<0.0001	16	34688.45	0.99	8777.47	0.93
Hennepin, MN	<0.0001	28	694.41	0.70	22344.01	0.93	<0.0001	28	377.75	0.84	9883.92	2.84
Huron, MI	<0.0001	2	26.67	0.31	736.73	6.16	No Abundance Data to Assess					
Ingham, MI	<0.0001	25	12.01	0.37	3864.95	14.42	No Abundance Data to Assess					
Lucas, OH	<0.0001	18	1273.15	0.91	7917.30	1.29	<0.0001	18	28398.61	1.00	11581.97	3.24
Macon, IL	<0.0001	2	857.04	0.11	60799.58	2.18	No Abundance Data to Assess					
Marion, IN	<0.0001	24	13797.02	0.99	639.78	0.28	<0.0001	29	1017.83	0.93	6975.74	1.21
Midland, MI	<0.0001	5	496.88	0.96	242.36	0.67	No Abundance Data to Assess					
Milwaukee, WI	<0.0001	16	5050.31	0.97	-1095.00	0.19	<0.0001	16	24846.45	0.99	2881.37	0.48
Monona, IA	<0.0001	2	66.99	0.23	2414.95	3.89	0.0002	2	9.23	0.11	479.85	1.52
Oakland, MI	<0.0001	13	623.89	0.96	270.74	0.35	No Abundance Data to Assess					
Polk, IA	<0.0001	26	730.56	0.68	23479.52	0.90	<0.0001	28	335.26	0.73	6915.36	0.65
Pottawattamie, IA	<0.0001	9	1479.42	0.96	551.90	0.41	<0.0001	20	71.73	0.36	-6687.77	0.06
Ramsey, MN	<0.0001	27	1320.89	0.84	16309.23	0.84	<0.0001	26	171.77	0.66	15564.91	6.99
Saginaw, MI	<0.0001	28	444.30	0.42	75371.97	2.21	No Abundance Data to Assess					
Scott, MN	<0.0001	27	484.89	0.76	13072.53	1.19	<0.0001	27	513.33	0.59	56141.66	4.62
St. Louis, MO	<0.0001	27	1213.59	0.55	91887.12	1.39	No Abundance Data to Assess					
Story, IA	<0.0001	27	469.53	0.83	5990.16	0.78	<0.0001	28	255.14	0.73	7372.34	0.98
Tuscola, MI	<0.0001	26	494.13	0.84	8744.95	1.45	No Abundance Data to Assess					
Washington, MN	<0.0001	27	139.83	0.61	7446.30	1.17	<0.0001	24	154.32	0.92	1247.76	1.52
Wayne, MI	0.0075	4	3.74	0.11	660.14	9.44	No Abundance Data to Assess					
Woodbury, IA	<0.0001	23	447.11	0.84	4219.36	0.71	<0.0001	23	197.60	0.85	283.18	0.28

Table 4.4. Estimates of cumulative monthly MIR and VI by each of four model scenarios: null, global, county-specific best fit, and county-specific global. Top cells indicate mean values, while bottom cells indicate cumulative (2000-2018) values for each respective outcome variable. Cells labeled no estimate (denoted *N.E.*) are a result of no or too few of mosquito abundance data to capture a reliable vector index value.

County	n	Null Model		Global Model		County-Specific Best Model		County-Specific Global Model	
		MIR	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*
Allen, IN	4739	2.48	9.21	3.76	10.11	2.77	10.11	2.77	10.11
		11738.18	43633.00	14352.85	38554.69	10545.73	38554.69	10545.73	38554.69
Anoka, MN	4996	11.61	6.15	9.24	5.67	11.70	5.83	11.65	5.83
		58013.18	30708.23	9.24	5.67	11.70	5.83	11.65	5.83
Bay, MI	2889	1.74	*N.E.*	*N.E.*	*N.E.*	1.88	*N.E.*	1.88	*N.E.*
		5038.61	*N.E.*	*N.E.*	*N.E.*	5038.61	*N.E.*	5038.61	*N.E.*
Carroll, IA	574	0.40	0.01	1.83	-0.63	0.26	0.01	0.20	*N.E.*
		230.12	5.72	1.83	-0.63	0.26	0.01	0.20	*N.E.*
Carver, MN	4167	6.36	3.16	7.00	4.24	5.92	4.24	6.20	4.24
		26503.18	13169.38	23141.33	2552.02	19129.12	2552.02	20477.50	2552.02
Champaign, IL	158	0.78	0.00	*N.E.*	*N.E.*	0.86	*N.E.*	0.86	*N.E.*
		123.69	0.00	*N.E.*	*N.E.*	123.69	*N.E.*	123.69	*N.E.*
Cook, IL	1583728	4.54	82.83	8.14	78.90	2.21	78.91	7.93	78.90
		7196235.02	131180756.90	1170826.13	1261882.79	35099.19	1261874.54	1141400.00	1261882.79
Dakota, MN	9194	7.60	7.09	6.63	8.84	7.60	8.84	7.62	8.84
		69911.08	65157.68	48430.00	11599.39	46553.86	11599.39	55715.56	11599.39
Dane, WI	1679	0.21	0.05	-0.70	0.07	0.27	0.07	0.24	0.07
		345.63	82.30	-446.71	37.75	306.31	37.75	152.91	37.75
Dubuque, IA	2216	0.17	0.00	1.53	0.01	0.96	0.01	0.14	*N.E.*
		380.76	3.03	317.67	0.40	729.01	0.40	28.94	*N.E.*
DuPage, IL	814	4.90	0.00	*N.E.*	*N.E.*	4.61	*N.E.*	4.63	*N.E.*
		3984.99	0.00	*N.E.*	*N.E.*	3670.48	*N.E.*	3689.56	*N.E.*
Hamilton, IN	4035	1.60	13.07	3.28	15.67	1.82	14.85	1.79	*N.E.*
		6436.54	52741.27	4230.45	30611.04	3554.01	48338.16	2304.33	*N.E.*
Hennepin, MN	12032	10.19	13.96	8.49	13.84	10.08	13.84	10.14	13.84
		122616.81	167991.32	79662.12	27686.55	83586.28	27686.55	95072.87	27686.55
Huron, MI	114	3.48	0.00	*N.E.*	*N.E.*	3.55	*N.E.*	0.78	*N.E.*
		397.09	0.00	*N.E.*	*N.E.*	404.77	*N.E.*	47.03	*N.E.*
Ingham, MI	564	5.17	*N.E.*	*N.E.*	*N.E.*	6.19	*N.E.*	6.19	*N.E.*
		2913.98	*N.E.*	*N.E.*	*N.E.*	2904.12	*N.E.*	2904.12	*N.E.*

Table 4.4. (continued)

County	n	Null Model		Global Model		County-Specific Best Model		County-Specific Global Model	
		MIR	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*
Lucas, OH	3376	16.59 56003.90	167.92 566908.60	11.84 26051.81	87.78 195836.10	16.73 39491.62	87.78 195845.77	17.70 38931.49	87.78 195845.86
Macon, IL	15043	1.71 25776.12	0.04 572.23	1.75 1473.07	13.77 6869.12	1.67 25183.61	*N.E.* *N.E.*	-0.07 -62.80	*N.E.* *N.E.*
Marion, IN	2976	3.62 10782.06	3.00 8917.71	3.47 375.22	3.97 8573.90	4.05 8374.27	3.97 8573.90	2.93 316.92	3.97 8573.90
Midland, MI	148	4.06 600.46	0.00 0.00	*N.E.* *N.E.*	*N.E.* *N.E.*	5.16 598.11	*N.E.* *N.E.*	5.24 607.70	*N.E.* *N.E.*
Milwaukee, WI	2744	0.80 2202.11	3.81 10442.47	5.63 4838.04	5.48 6690.63	1.01 2093.21	4.60 9580.35	0.98 846.84	*N.E.* *N.E.*
Monona, IA	434	2.60 1129.22	0.13 57.01	4.69 131.22	1.13 112.17	2.60 1129.22	-0.18 -68.70	0.81 22.82	*N.E.* *N.E.*
Oakland, MI	424	2.41 1021.80	*N.E.* *N.E.*	*N.E.* *N.E.*	*N.E.* *N.E.*	1.91 658.17	*N.E.* *N.E.*	1.91 658.17	*N.E.* *N.E.*
Polk, IA	19281	2.48 47899.05	0.43 8328.07	3.25 18513.24	0.78 2718.49	1.24 11248.98	0.84 2919.69	1.05 5964.55	0.84 2919.69
Pottawattamie, IA	4323	1.79 7756.93	0.02 74.06	0.85 198.76	0.07 40.21	2.53 1593.59	0.03 66.58	-1.36 -319.93	0.07 40.21
Ramsey, MN	10101	12.36 124835.26	18.08 182577.75	12.24 88411.92	21.01 48506.92	12.08 80824.39	21.01 48506.92	12.20 88105.80	21.01 48506.92
Saginaw, MI	19084	1.24 23714.18	*N.E.* *N.E.*	*N.E.* *N.E.*	*N.E.* *N.E.*	1.37 23303.93	*N.E.* *N.E.*	*N.E.* *N.E.*	*N.E.* *N.E.*
Scott, MN	11704	7.91 92582.68	5.30 62010.82	6.79 41313.03	8.54 13756.66	8.34 34114.05	6.36 60461.23	7.85 47715.84	8.54 13756.66
St. Louis, MO	29338	2.39 70012.49	*N.E.* *N.E.*	*N.E.* *N.E.*	*N.E.* *N.E.*	2.56 67374.59	*N.E.* *N.E.*	*N.E.* *N.E.*	*N.E.* *N.E.*
Story, IA	11850	1.47 17381.19	0.39 4597.38	1.73 4597.33	0.98 2597.08	1.13 2999.55	0.76 2013.71	1.14 3008.64	0.75 1990.41
Tuscola, MI	2823	2.09 5893.31	*N.E.* *N.E.*	*N.E.* *N.E.*	*N.E.* *N.E.*	2.37 5765.01	*N.E.* *N.E.*	2.37 5765.01	*N.E.* *N.E.*
Washington, MN	3283	9.76 32029.54	2.61 8567.29	5.34 13951.36	3.86 1290.54	9.57 22481.71	4.40 1468.29	9.78 25562.24	4.38 1461.37
Wayne, MI	89	5.82 517.80	1.17 104.24	*N.E.* *N.E.*	*N.E.* *N.E.*	5.82 517.80	*N.E.* *N.E.*	8.26 421.02	*N.E.* *N.E.*
Woodbury, IA	6789	1.05 7138.29	0.14 973.57	0.88 1342.72	0.23 193.09	0.89 1884.64	0.32 266.34	0.67 1024.79	0.32 266.42

Table 4.5. Assessment of frequency and strength of covariates used by each county's best-fit global MIR and VI models. Strength of covariates determined by LogWorth (-log10(p-value)), resulting from standard least squares regressions. Any value greater than 1.3 is statistically significant ($p < 0.05$). Values in gray indicate summary values. With the exception of % high school and % bachelor's degree, no other socio-economic, demographic, or land-use/land-cover variables were included in any county's final model.

	County-Specific Best-Model	**Allen, IN**		**Anoka, MN**		**Bay, MI**		**Carroll, IA**		**Carver, MN**		**Champaign, IL**		**Cook, IL**	
	Dependent variable	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*
Independent Variable	Year	9.57	1.19	7.49	4.79	42.76			91.24	238.23				61.76	58.59
	month			211.07	28.20	151.30			39.12	68.11	10.39			0.00	369.16
	# collected	2.28		1.48											
	#tested		0.42		3.81	0.23					2.30			11.99	16.22
	WNV Positive Mosquito			27.18	0.94	8.77				6.95				112.28	161.41
	mirlag1	94.85		37.29						33.37				1928.94	
	VIlag1		22.37		9.24						10.09				102.56
	Jantmean			4.72	0.59	7.50				8.06				0.00	76.85
	Jantmin			4.73	0.59	7.51			59.62	8.04	8.26			0.00	76.90
	Jantmax			4.70	0.59	7.50		1.00		8.08	8.56	15.42		0.00	76.81
	minppt	4.54	35.77	0.48	2.23	1.33		19.30	97.97	36.51	5.87			43.37	90.13
	maxppt	16.89	69.38		3.91	37.17				110.81	2.51				125.44
	meanppt			1.42	0.15	23.29					0.82				
	sumppt									76.76				34.45	126.97
	mintd	26.34	9.40	14.74	2.22	3.69				19.04	3.75			0.00	68.23
	maxtd	2.49	22.60	31.71	0.47	6.99				10.55	1.20			0.00	65.31
	meantd			11.59	1.67						5.00			0.78	68.47
	sumtd					7.43				21.30					
	mintmax	65.40		26.95	4.46	8.71				73.77	2.51			0.00	34.83
	maxtmax	21.90	30.66	15.91	1.23	67.02				105.15	1.74			38.74	1.22
	meantmax				11.92	18.02				49.51				0.00	
	sumtmax			61.04							1.56			0.00	1.17
	mintmean	0.35		7.96	2.70	50.78				15.54	3.44			14.02	10.34
	maxtmean	5.56		37.85		49.51				97.97				4.93	1.00
	meantmean			61.35		16.66				50.10		2.67			
	sumtmean				9.13						1.56			18.63	1.16
	mintmin	7.39	0.42	54.67	12.08	2.55				12.13	4.20	2.06		42.45	4.65
	maxtmin	6.02		26.29	0.65	11.71				78.75	4.52				1.98
	meantmin					17.06								18.69	
	sumtmin			60.54						50.11	1.51				1.15
	meantmeanlag1	85.72	84.26	51.06	0.28	33.53			17.43	9.83	2.89	3.42		19.24	16.53
	meanpptlag1	2.06	6.86	123.30	0.37	49.68			24.77	8.21	1.47	1.55		25.12	68.38
	%_highschool_deg														
	%_bachelor_deg														
	n covariates	15.00	11.00	24.00	23.00	24.00	0	2.00	6.00	24.00	21.00	5.00	0	24.00	25.00
	sum LogWorth	351.34	283.34	885.52	102.21	630.69	0	20.30	330.15	1196.86	84.15	25.12	0	2375.40	1625.44
	mean LogWorth	23.42	25.76	36.90	4.44	26.28	N/A	10.15	55.02	49.87	4.01	5.02	N/A	98.97	65.02

Table 4.5. (continued)

	County-Specific Best-Model	Dakota, MN		Dane, WI		Dubuque, IA		DuPage, IL		Hamilton, IN		Hennepin, MN		Huron, MI	
	Dependent variable	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*
Independent Variable	Year	17.83	11.97	17.41	0.01	2.54	83.78	17.69		1403.38	447.08	1.92	1.47	9.92	
	month	7.39	41.49	235.88	30.16	0.54	119.52	223.63		1758.09	979.66	162.53	52.13		
	# collected	0.99										1.77			
	#tested		3.65					0.88					9.20	0.47	
	WNV Positive Mosquito	28.97	3.84		6.53		0.37			0.26		30.57	1.40		
	mirlag1	24.89		45.19						283.45		524.02			
	VIlag1		41.90		8.48		90.26				610.78		16.35		
	Jantmean	44.32	38.28	63.32	11.38	5.52				1309.01		0.85	2.72		
	Jantmin	44.38	38.28	63.33	11.39	4.55	76.81			1312.84	681.96	0.86	2.72		
	Jantmax	44.22	38.28	63.30	11.38						667.13	0.83	2.72		
	minppt	3.62	6.30	141.37	31.41	5.22	68.27			1122.48	571.41	85.59	41.93		
	maxppt	8.66	2.69	19.09	3.23	18.43		173.42		1114.41		75.79	41.21		
	meanppt	10.23		83.29		2.89				1231.31	342.00	82.69			
	sumppt				20.58		72.14						47.17		
	mintd	44.49	4.81	18.68	19.86					6.64	660.70	6.89	26.20		
	maxtd	11.10	4.15	80.65	10.66		67.87			706.48		86.14	97.33		
	meantd			9.88									24.10		
	sumtd	28.26	3.64				80.48			237.90		41.56			
	mintmax	9.45	20.41	86.81	7.25						235.52	8.37	2.25		
	maxtmax	1.43	5.66	4.04	11.14						139.01	1.77	48.54		
	meantmax				5.86						20.71				
	sumtmax	9.97	5.69	83.09								82.23	5.96		
	mintmean		2.13	130.80	14.27						61.22		2.29		
	maxtmean	1.10	22.07	13.60	0.04							5.79	46.13		
	meantmean				5.69								6.08		
	sumtmean	10.12	5.93	82.82		9.03						82.29			
	mintmin		0.18	77.03	11.93							101.29	55.18		
	maxtmin	0.02	11.98	77.67	3.80		106.02			189.74	6.36	16.16	72.49		
	meantmin	10.12	5.99				108.13								
	sumtmin			83.43	5.62							82.88	5.96		
	meantmeanlag1	5.93	16.78	117.23	8.93		73.73	63.61		830.71	984.85	29.65	42.48		
	meanpptlag1	91.79	62.99	11.17	13.40	0.67	109.38	58.86		11.12	6.29	47.99	14.08		
	%_highschool_deg														
	%_bachelor_deg														
	n covariates	23.00	24.00	23.00	23.00	9.00	13.00	6.00	0	15.00	15.00	24.00	25.00	2.00	0
	sum LogWorth	459.26	399.10	1609.07	252.98	49.38	1056.76	538.08	0	11517.81	6414.67	1560.42	668.06	10.39	0
	mean LogWorth	19.97	16.63	69.96	11.00	5.49	81.29	89.68	N/A	767.85	427.64	65.02	26.72	5.20	N/A

Table 4.5. (continued)

County-Specific Best-Model	Ingham, MI		Lucas, OH		Macon, IL		Marion, IN		Midland, MI		Milwaukee, WI		Monona, IA	
Dependent variable	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*
Year	2.52		408.36	138.67			418.22	5.53	14.26			3.53		
month	5.33		153.60				870.66	275.70			973.34	977.18		
# collected														
#tested									2.19					
WNV Positive Mosquito	1.44							1.09						
mirlag1							399.32				703.45			
VIlag1				529.31				75.40				481.04		
Jantmean	2.43		17.65		32.27			17.40				4.56		
Jantmin	2.43		17.68				466.38	17.40			12.43	4.62		
Jantmax	2.43		17.62		81.17		125.13	17.40			8.74	4.46		
minppt	1.42			721.81			380.34	10.48				458.21		
maxppt	5.99		49.49	344.87			611.81	1.23			180.42	657.97		
meanppt				755.15										
sumppt	5.28		28.27				487.11	17.81			266.13	590.57		
mintd	2.94		39.92	0.45			206.14	5.72			16.29			4.20
maxtd	0.19		179.05	56.26			4.44	44.26						
meantd	4.04		141.86	43.48			55.64							
sumtd								25.19						
mintmax	3.51		95.89	586.07				89.51	4.56					
maxtmax	2.67		24.79	596.14			3.41	28.10					13.57	
meantmax	2.90							3.99			6.33		15.73	
sumtmax				10.54			8.43		1.02					
mintmean	4.05						138.74	27.62				308.65		
maxtmean				712.98			60.25	21.23						
meantmean	2.90						0.00							
sumtmean				520.86				3.99			2.12			
mintmin	6.64		3.88				616.65	6.71			100.07	496.16		
maxtmin								0.10			28.45			
meantmin	2.90		13.82					3.85						
sumtmin							9.73					429.13		
meantmeanlag1	2.71		8.20	555.88			410.06	21.41	20.77		861.91	24.87		
meanpptlag1	4.71			856.77			0.83	46.98			198.09			1.82
%_highschool_deg														
%_bachelor_deg								14.67						
n covariates	21.00	0	15.00	15.00	2.00	0	20.00	25.00	5.00	0	13.00	13.00	2.00	2.00
sum LogWorth	69.43	0	1200.07	6429.25	113.44	0	5273.26	782.77	42.80	0	3357.74	4440.92	29.29	6.01
mean LogWorth	3.31	N/A	80.00	428.62	56.72	N/A	263.66	31.31	8.56	N/A	258.29	341.61	14.65	3.01

Independent Variable

Table 4.5. (continued)

Independent Variable	County-Specific Best-Model	Oakland, MI		Polk, IA		Pottawattamie, IA		Ramsey, MN		Saginaw, MI		Scott, MN		St. Louis, MO	
	Dependent variable	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*
	Year			123.20	41.17	9.18	66.03	406.67	7.78	1.42		209.76	23.04	677.57	
	month	30.76		161.04	113.77	202.41	44.43	11.62	84.27	0.00		87.14	55.81	121.48	
	# collected							2.93				1.36			
	#tested	0.84			2.49				11.30	2.52				4.31	
	WNV Positive Mosquito				16.46			15.58	10.56	66.16		5.64		246.90	
	mirlag1			159.89		24.30		5.34				97.38		352.55	
	VIlag1				269.44		7.51		2.78				435.18		
	Jantmean			14.58	34.01	101.96		0.60		0.00		3.01	0.21	98.06	
	Jantmin	1.84		14.53	34.05	102.13		0.58		0.00		3.01	0.21	98.08	
	Jantmax	1.79		14.62	33.98	101.80	25.30	0.63	5.98	0.00		3.01	0.21	98.02	
	minppt			98.65	0.58	93.08	1.25	16.55	0.55	0.96		10.22	28.05	66.82	
	maxppt			110.96	2.07	27.12	1.78	2.41	7.57	63.14			0.90	191.36	
	meanppt						1.02						11.06	173.16	
	sumppt			192.04	0.15				2.28	38.26		5.73			
	mintd	1.75		115.60	4.86		41.65	188.73	21.59	69.76		96.75	32.90	183.35	
	maxtd			1.93	25.25			2.82	18.02	220.11		79.65	15.86	58.63	
	meantd								22.66			100.62	26.50	17.74	
	sumtd			87.98	23.31		29.07	99.86		132.45					
	mintmax	0.34		418.67	12.32		6.20	150.63	1.27	0.00		1.28	37.38	249.35	
	maxtmax			22.94	9.71		0.19	2.85	8.07	3.53		9.52	16.23	738.91	
	meantmax								39.06	0.00			0.72		
	sumtmax	0.17		54.95	22.81		0.81	18.92		0.00		1.14			
	mintmean			345.00	13.85			332.87	1.19	15.69		11.67	0.60	56.90	
	maxtmean	0.16		124.14	18.54		2.30	3.07	3.77	16.92		73.13	116.51	5.32	
	meantmean			54.57	23.03										
	sumtmean							19.15	38.88	32.66		1.18	0.67	486.46	
	mintmin			4.02	5.90			80.59	32.96	12.39		1.58	0.25	43.49	
	maxtmin	0.75		12.29	4.63		5.46	67.01	90.82	30.50		10.28	189.29	4.02	
	meantmin			53.50	22.81			18.21							
	sumtmin								39.94	33.25		1.16	0.64	204.52	
	meantmeanlag1	12.06		89.82	12.13		5.89	79.40	0.85	41.02		59.82	0.59	1.62	
	meanpptlag1	1.19		120.73	5.47	39.29	19.41	10.25	25.36	0.25		2.26	224.94	114.53	
	%_highschool_deg												37.35		
	%_bachelor_deg														
	n covariates	11.00	0	23.00	25.00	9.00	16.00	24.00	23.00	25.00	0	24.00	24.00	24	0
	sum LogWorth	51.65	0	2395.61	752.78	701.27	258.28	1537.24	477.49	781.01	0	876.31	1255.10	4293	0
	mean LogWorth	4.70	N/A	104.16	30.11	77.92	16.14	64.05	20.76	31.24	N/A	36.51	52.30	178.88	N/A

Table 4.5. (continued)

Independent Variable	County-Specific Best-Model	Story, IA		Tuscola, MI		Washington, MN		Wayne, MI		Woodbury, IA				
	Dependent variable	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	*MIR*	*VI*	n counties	sum LogWorth	mean LogWorth
	Year	73.00	39.88	234.33		13.22	6.30			261.64	0.26	45	5716.15	127.03
	month	127.81	162.96	95.39		7.54	1.27	2.38		58.73	26.98	44	9012.64	204.83
	# collected	1.91				5.25						8	20.78	2.60
	#tested		2.70	5.21			3.15				1.92	20	104.95	5.25
	WNV Positive Mosquito		21.85	14.91		21.40	1.39				8.08	26	816.51	31.40
	mirlag1	32.34				58.76				68.70		18	4739.94	263.33
	VIlag1		69.50				5.72				140.35	19	2947.24	155.12
	Jantmean	45.08	64.43	0.95		16.64	5.80				20.91	33	1994.83	60.45
	Jantmin	45.09	64.46	0.96		16.65	5.80				20.92	39	3277.25	84.03
	Jantmax	45.06	64.40	0.94		16.64	5.81			42.31	20.91	41	1636.34	39.91
	minppt	26.27	10.43	103.71		10.95	36.00			5.78	6.89	43	4511.25	104.91
	maxppt	40.76	100.88	76.29		26.11	36.86	0.13		53.23	1.92	40	4386.00	109.65
	meanppt			111.99						14.38		16	2850.60	178.16
	sumppt	0.95	7.29			9.66	38.22					21	2082.59	99.17
	mintd	6.30	20.36	92.41		4.18	4.71			122.22		40	2176.61	54.42
	maxtd	124.49	84.16	3.38		2.76				182.06	23.29	36	2230.04	61.95
	meantd		58.88			3.47	9.94			184.98		19	806.24	42.44
	sumtd	49.70		26.15							22.07	16	854.36	53.40
	mintmax	28.72	39.21	50.86		43.37				11.09	41.58	36	2387.07	66.31
	maxtmax	13.81	40.87	102.61		9.68	22.45	2.54		4.68	30.40	39	2202.02	56.46
	meantmax					4.97	16.01			45.32		16	254.15	15.88
	sumtmax	30.57	25.02	4.59							29.63	23	441.61	19.20
	mintmean	4.36	0.44	3.98		1.74	5.31			6.91	63.05	32	1681.70	52.55
	maxtmean	10.11	60.87	34.42		2.94	10.02			10.63	31.74	32	1619.66	50.61
	meantmean			4.75		4.86	15.97			44.88	35.56	15	341.17	22.74
	sumtmean	30.00	24.65									20	1361.17	68.06
	mintmin	36.86	13.64	4.02			0.14			14.46	9.28	35	1861.98	53.20
	maxtmin	0.46	26.10	5.35		1.92	23.07			7.76	11.44	35	1161.60	33.19
	meantmin		24.78	4.84			15.93			44.27		15	366.89	24.46
	sumtmin	30.01				4.84						17	1031.40	60.67
	meantmeanlag1	3.72	5.43	23.66		6.22				64.90		44	4744.87	107.84
	meanpptlag1	9.44	6.03	9.20		12.70	0.59			95.67		45	2481.80	55.15
	%_highschool_deg											1	38.35	38.35
	%_bachelor_deg											1	15.67	15.67
	n covariates	24.00	25.00	24.00	0	24.00	22.00	3.00	0	21.00	20.00	66	902.00	13.67
	sum LogWorth	816.81	1039.23	1014.89	0	306.48	270.43	5.05	0	1344.61	547.16	66	71233.49	1079.30
	mean LogWorth	34.03	41.57	42.29	N/A	12.77	12.29	1.68	N/A	64.03	27.36	54	4302.40	79.67

FIGURE 4.1. Distribution of mosquito abundance and submitted infection data, aggregated by county, in the midwestern United States. Presence of infection in mosquito pools are indicated by color. Major cities located with the study area are indicated by green dots.

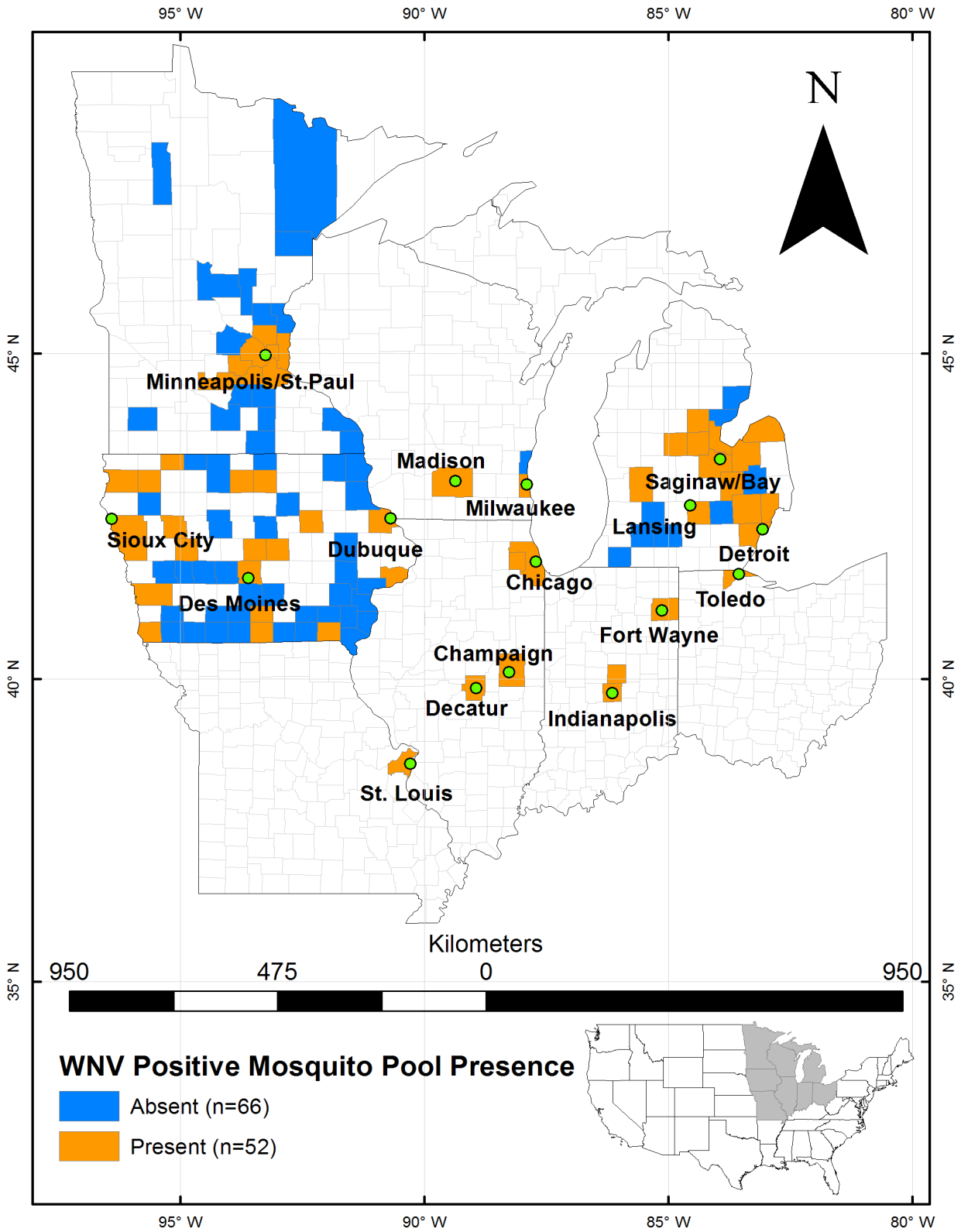


FIGURE 4.2. Predicted WNV presence, modeled from 2000-2018 cumulative infection data by county. Dependent variables, minimum infection rate (MIR, A1-D1) and vector index (VI, A2-D2), are predicted under four model conditions: null (A), global (B), county-specific global (C), and county-specific best-fit (D). Predicted values indicate WNV mosquito infection by quintiles.

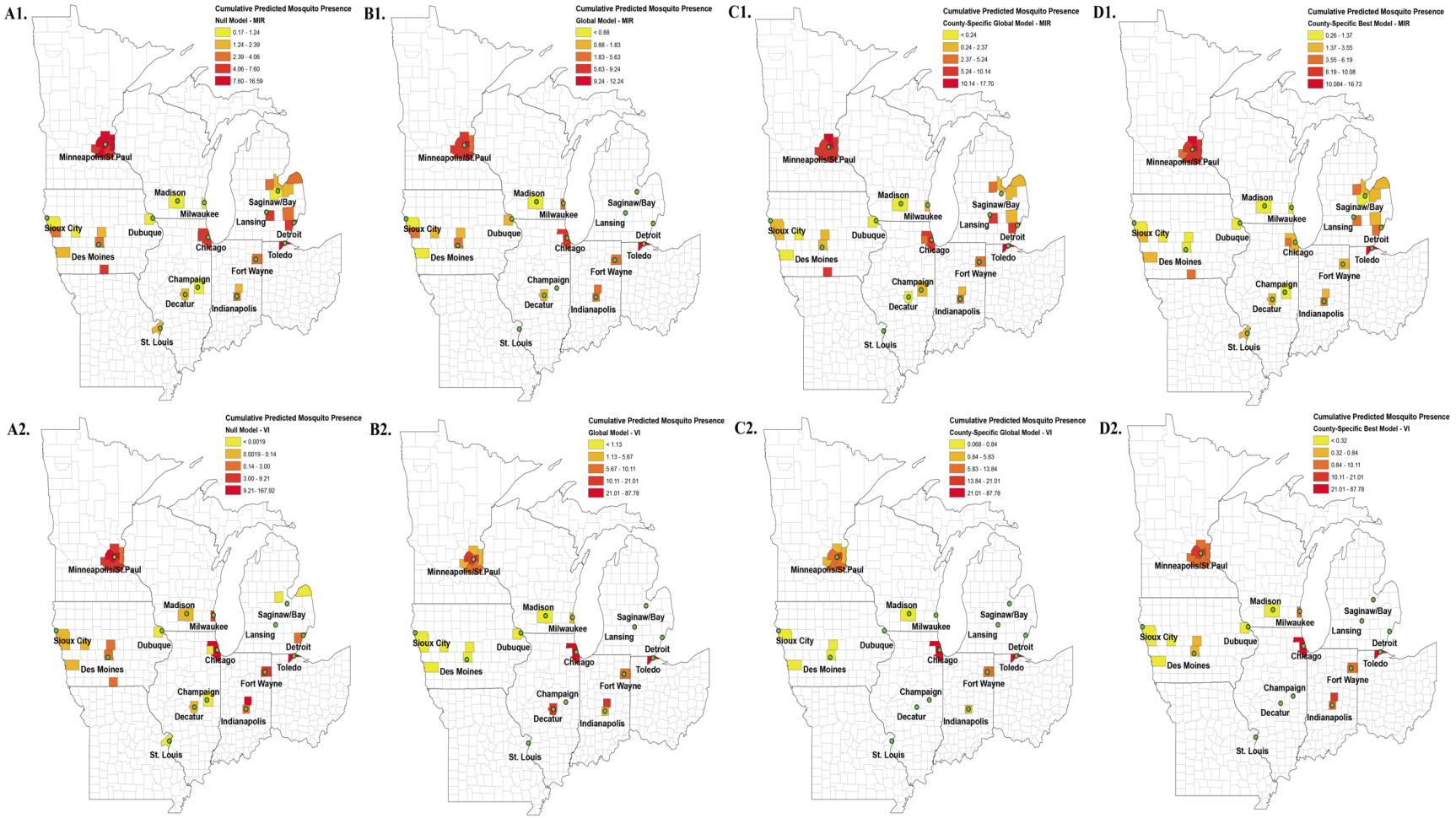


FIGURE 4.3. Prediction of MIR (A1-D1) and VI (A2-D2) for the entire study area by inverse distance weighting (IDW) interpolation methods. Values determined under four model conditions: null (A), global (B), county-specific global (C), and county-specific best-fit (D), are indicated by n=8 natural breaks in the distribution of predicted mosquito infection values.

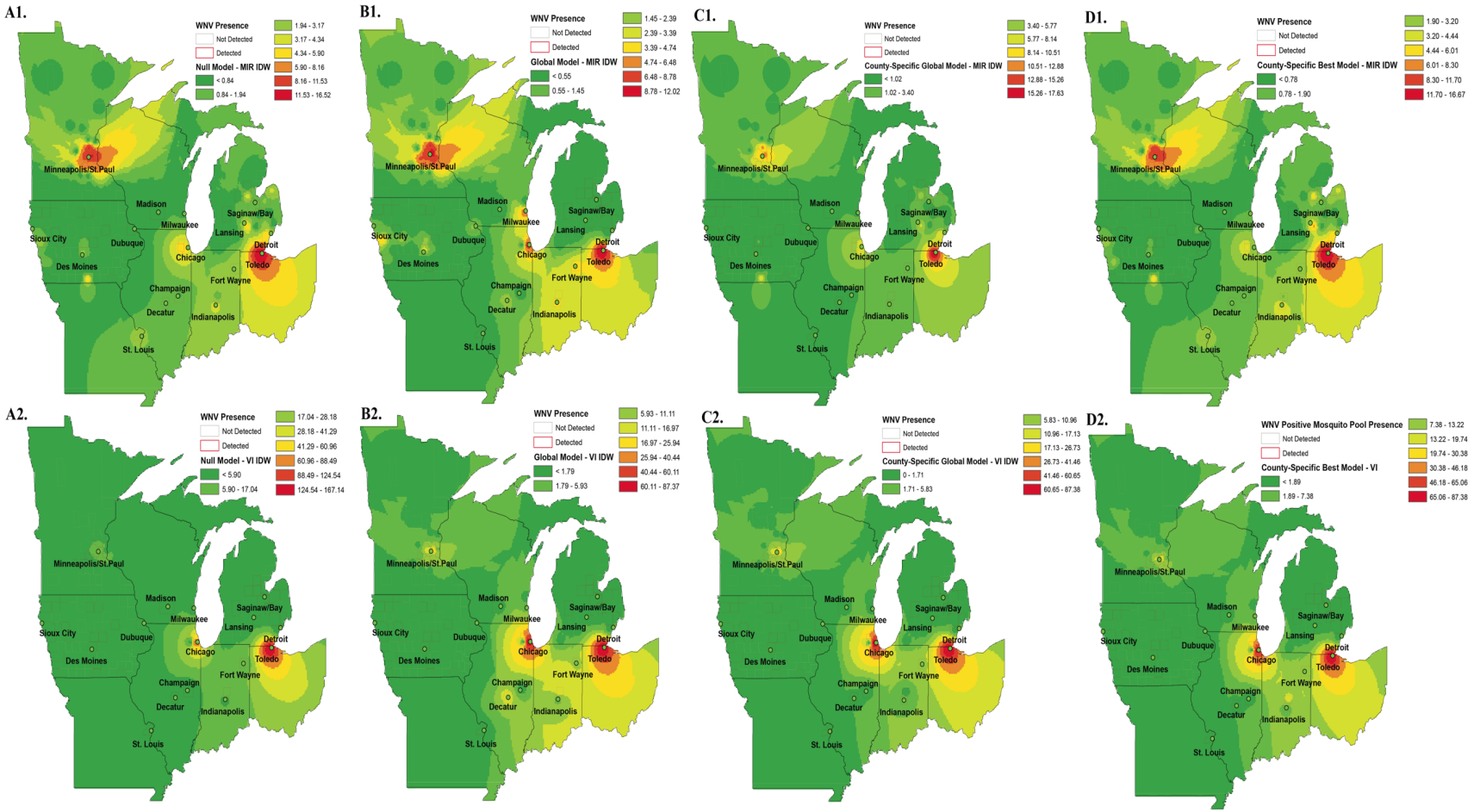


FIGURE 4.4. Human WNV neuroinvasive illness in the midwestern United States, from 2000-2018. Cumulative (A) and infection rates (per 100,000 people, B) are indicated by quintiles of infection per county.

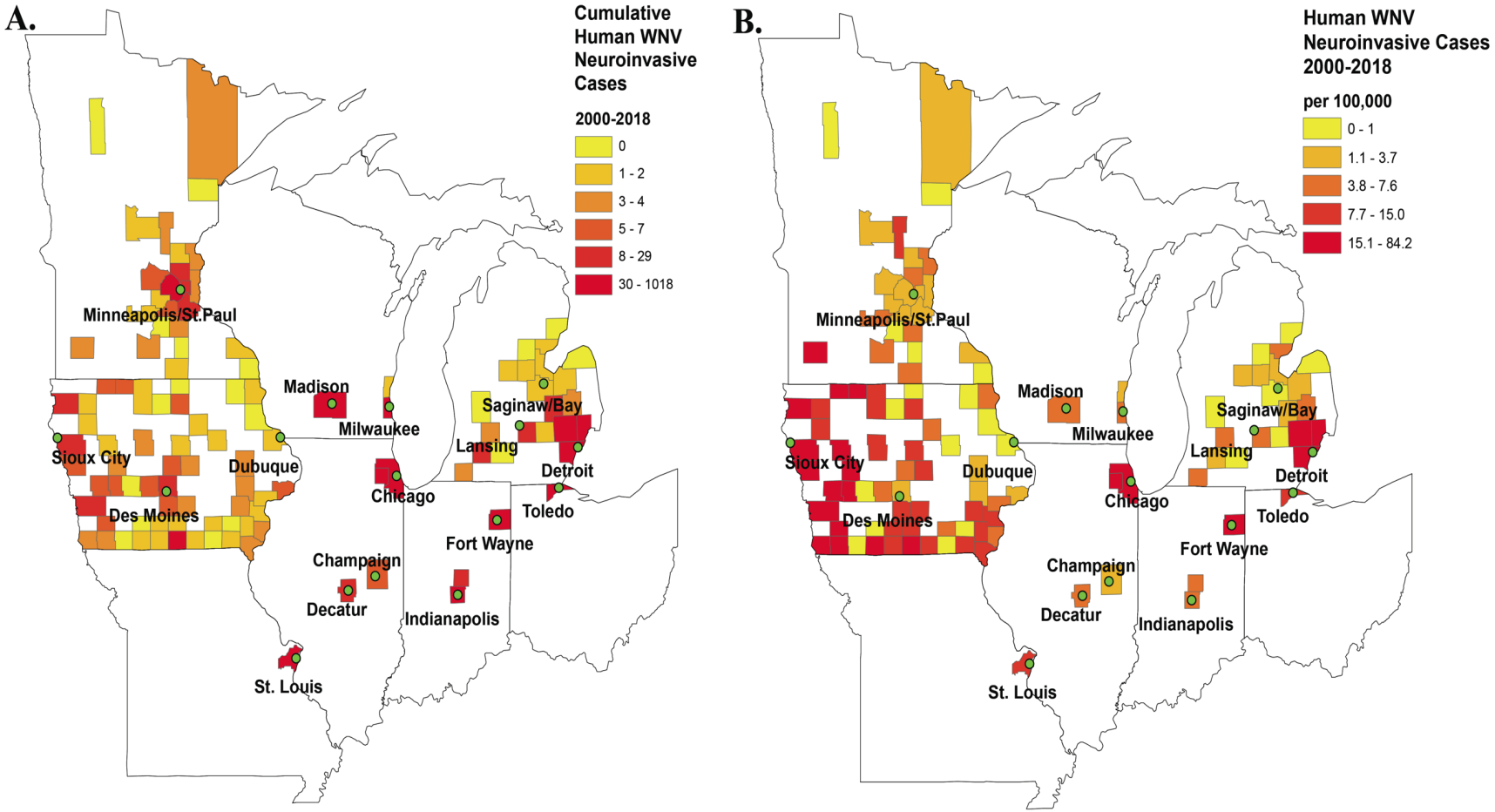
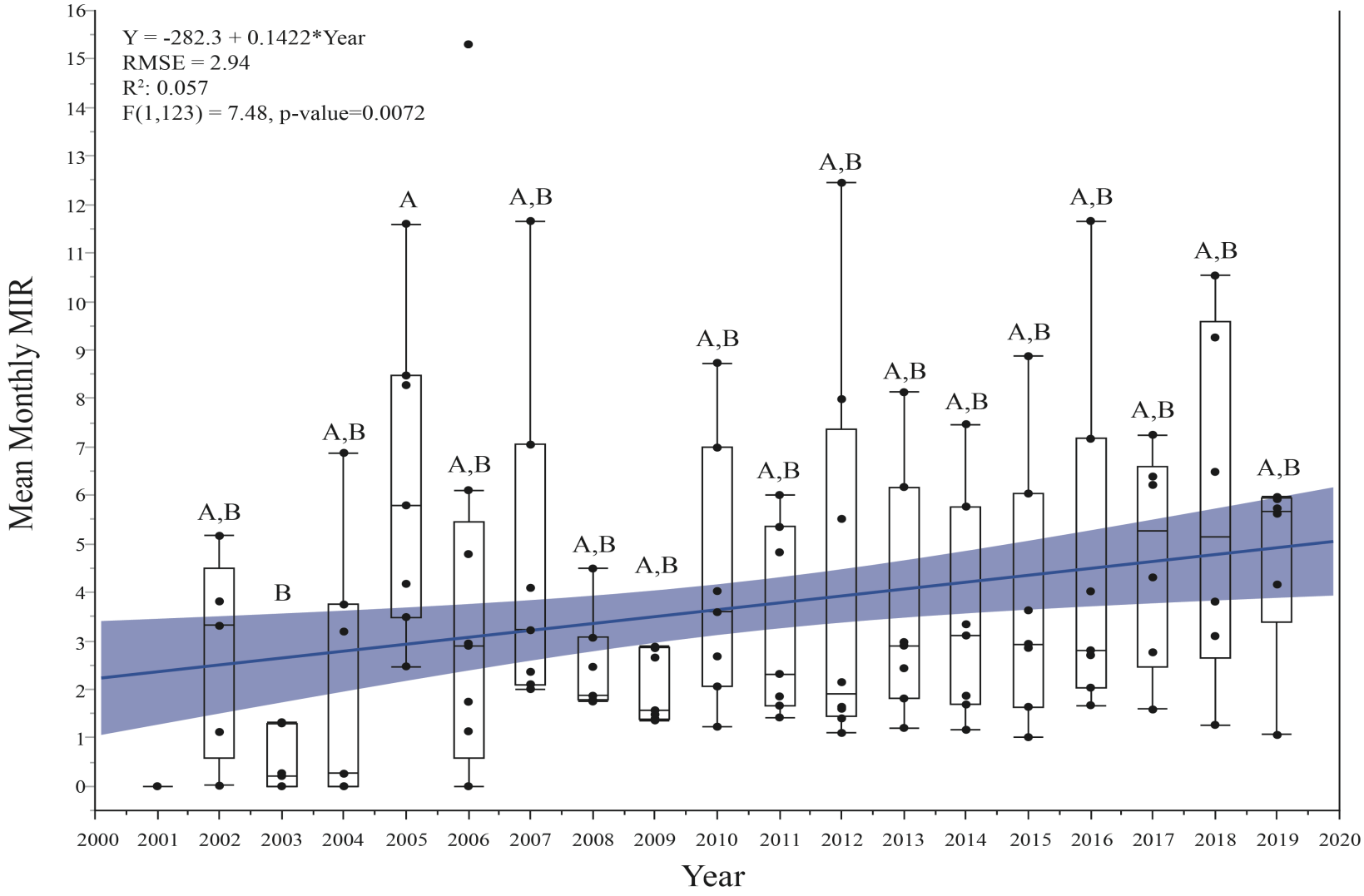


FIGURE 4.5. Annual mean MIR derived from 118 counties in the midwestern U.S. from 2000-2018. Letters above each respective bar value are designated classification values for mean differences by Tukey's HSD. Error bars denote 1 standard error.



CHAPTER 5: CONCLUSIONS, RECOMMENDATIONS, AND FUTURE DIRECTIONS

5.1. KEY MESSAGES

5.1.1. Chapter 2

1. The factors and their overall effect on the prediction of human WNV cases differs across scale. Although improved, in comparison to the control Cook/DuPage model applied to the same study region, the "best fit" UFS model AUC = 0.89, suggesting newly unaccounted variances are present.
2. Both VI and MIR contribute to high performing human WNV prediction models under UFS study areas. In direct comparison, VI is favorable to MIR. However, given limited resources in acquiring and processing additional data, MIR is more efficient for predicting human WNV illness.
3. The effort and resources required to acquire additional covariates, most of which are not publicly available, demonstrate a slight improvement in model prediction and appear less important in reducing variance.
4. In addition to the conventional WNV covariates, namely weather and infection rates, land-use and land-cover and SES/demographic information are widely available with little to no processing or analyses required, and provide the breadth to develop excellent prediction models. However, any covariate utilized should be structured at the finest spatial and/or temporal resolution possible.
5. Human exposure to mosquito biting rates provided minimal benefits to model prediction. However, results from these efforts provided potentially key insight to the susceptibility of humans in locations where WNV is prevalent. Additionally, where WNV is less of a

concern, these results provide insight into nuisance mosquito exposure that may lead to improvements in targeted control efforts.

5.1.2. Chapter 3

1. Evaluating ecological processes all too often occurs at scales larger than the process of the underlying mechanisms. This study focused on addressing this problem, and compared overall performance of best-fit models and covariates from three scales in the Chicago, IL region.
2. Overall, as extent increased, model performance increased (RMSE decreased, ROC increased). Additionally, RMSE was significantly higher only when assessing UFS model performances. Our findings suggest that as extent decreases, covariate importance increases.
3. As scales increase and the sampling of locations decrease, the number of observations decrease, and the risk of not capturing heterogeneity that is representative of the entire study area is high. This phenomenon may have occurred at the UFS, as indicated by higher performing models fit to smaller scales.
4. Of the 23 covariates assessed in this study, only 6 (dlipct, Jantemp, MIRlag4, templag3, templag4, totpop) were included across multiple scales. Despite this variation, fitting one scale's best-fit model to another resulted in marginally weaker, but still high performing models. While these variances are mostly captured, the small, but subtle differences in results elude to differences in underlying ecological mechanisms of WNV that behave uniquely at different scales.
5. These small changes are highlighted by translating each of the scales' best-fit models into human WNV risk maps. Overall, these maps display similar distributions of disease risk

(in overlapping regions), but vary in frequency and size, increasing in patchiness at finer-scales.

6. Despite these subtle differences, these models and risk maps provide useful visual and statistical inferences into the ecological processes of WNV across scales. These tools may be particularly helpful for public health, mosquito, and disease control personnel in predicting and preventing disease before zoonotic spillover occurs.

5.1.3. Chapter 4

1. An 18-year analysis of WNV in the midwestern United States resulted in a total of 6,610,507 collected and 4,710,953 tested female *Culex* mosquitoes from 118 counties in 8 states.
2. Mosquito infection, provided as MIR and VI, was evaluated under four model scenarios. Overall, each model predicted MIR and VI values that were similar to their counterparts, with the majority of final county values derived from well-fit estimates ($R^2 > 0.85$, RMSE < 1.5).
3. Counties with the highest resources and greatest surveillance efforts generally do not translate to higher mosquito infection rates.
4. When compared to human neuroinvasive illness, mosquito infection models adequately represent where human illness is highest. However, results from this study suggest many missing knowledge gaps in environmental viral prevalence among rural counties, particularly throughout Iowa and southern Wisconsin.
5. WNV infected mosquitoes are increasing in the Midwest by an estimated 14.2% annually. A follow-up manuscript will evaluate these trends with projected changes in climatic and human-derived forces (e.g. population growth, land-use/land-cover).

5.2. CONCLUSIONS

This dissertation evaluated the relationships of human and mosquito WNV infection to a wide variety of potential explanatory variables. These explanatory variables, all related to the ecology of WNV, encompass elements related to the biology of the virus, past climate and weather, environment, and human-derived. The collection of data related to WNV was a massive feat, but resulted in the use of new and rarely used information that provided new insights into potential routes of exposure. The greatest example of this is the collection and usage of human and mosquito behavior data. Using human landing catches and observing human activity in public spaces, this dissertation was able to generate two indices: the human WNV risk and nuisance factor indices. The human WNV risk indices is derived from the measured exposure (a factor of observed age, length of activity, and type of activity) of human beings combined with the results of the mosquito landing rates, recorded in the same locations. Additional analyses are ongoing with these data, but preliminary results demonstrate an association linking high *Culex* abundance with the most human exposure in high risk WNV locations. This finding lays out a potentially dangerous combination of factors that set the stage for human transmission. Examples like this are strategies future research and surveillance efforts should emphasize and incorporate.

Surprisingly, linear regression models for nearly all locations across each study were high performing. Overall, this dissertation demonstrated that the collection of high resolution spatiotemporal data does improve model performance. However, not all covariates are equally meaningful, as some proved very time-consuming to collect, process, and analyze, with only a slight improvement in reducing model variance. These results led to a new discussion regarding the trade-off between resource allocation and benefit in understanding WNV transmission. Our study attempted to quantify this tradeoff, and suggested covariates that may be more important to

acquire given limited resources.

This dissertation also highlights the effects of scale in WNV disease ecology. We found that when evaluating WNV under an increasing spatial extent, overall model strength increases. Not surprisingly, this phenomena is a common occurrence in ecological studies, but highlights a critical finding: these methods, traditionally used in numerous past studies, often do not capture the fine-scale ecological processes of WNV transmission, and are likely the most important information detailing how human infection occurs. Flipping this around, evaluating models at very fine scales results in poorer performing models, as compared to broad scales, but suggest fine-scale ecological processes are overlooked.

Lastly, this dissertation attempted an ambitious evaluation of all known mosquito infection in the midwestern U.S. from 2000-2018. Motivated by a lack in standardized evaluation of WNV, especially across regions, this study collected, processed, and analyzed over 6.6 million records of *Culex* abundance and 4.7 million records of *Culex* infection data. Our results indicated that surveillance and testing efforts are representative of human neuroinvasive cases by county, except in nearly all of Iowa and southern Wisconsin. In these counties, reported mosquito abundance and infection are not accurately representing human risk, indicating a general lack of knowledge in these areas. We suggest that an emphasis in surveillance for these counties will result in a better understanding of mosquito infection, providing necessary mosquito control when necessary. Additionally, this study found that locations with the most resources to combat WNV do not necessarily find the most infection. Several rural counties, albeit with less historical data available, may provide a higher WNV risk than urban counties in the Midwest. However, the overall 18-year history of WNV shows three distinct hot-spots of mosquito infection: Minneapolis/St.Paul, Chicago, and Toledo/Detroit. Lastly, this study found that mosquito

infection is increasing in the Midwest by about 14.2% and is likely not an artifact of increased surveillance.

While conducting human landing rates in the field, we found that *Cx. salinarius* was by far, the most abundant *Culex* species captured. Contrary to expectation, our collections only resulted in the capture of a single *Cx. restuans* (which was a male) and no *Cx. pipiens*. Known as the unbanded saltmarsh mosquito, *Cx. salinarius* is a highly competent bridge vector, with a high affinity for mammalian blood. Previous studies have indicated that over 50% of *Cx. salinarius* blood meal contents are mammalian (Molaei et al. 2006). Most commonly distributed throughout brackish and marshy environments in states along the Gulf of Mexico, Georgia and the Carolinas, this mosquito appears to be abundant throughout the northwest suburbs of Chicago. Although not reported in this dissertation, but as part of other related projects stemming from this work, these results provide potentially critical information in the transmission of WNV in humans in Chicago. Previous efforts by Dr. Surendra Karki and Dr. Marilyn O'Hara Ruiz have provided this research with the evidence to investigate fine-scale transmission in specific locations, several of which had high human infection, but were not being captured by models.

This dissertation, along with several previous studies, finds that WNV continues to prevail in the environment, despite active control and surveillance efforts. However, through our efforts to improve model predictions, we found two alarming findings that will likely have profound impacts on future infection in the region: the annual increase in mosquito infection and the abundance of *Cx. salinarius.* Findings from this research also suggest that the increase in mosquito infection is not affected by resources or capabilities of local mosquito control agencies. Translating to human health, the past three outbreaks are an ominous warning that future outbreaks are highly likely, and will continue if infection persists in mosquitoes and avian hosts.

As a future post-doctoral associate in the Smith lab, I will conduct future studies that stem from this research with new methods in disease forecasting. In particular, my work will apply the models generated in the Chicago and Midwest towards future local, regional, and national climate predictions. The main motivation for these future studies is to provide reliable information that can be used by decision-makers to help prepare for varying levels of risk given the constantly changing dynamics that the climate, environment, and humans have on disease ecology. Personally, it is my goal that the methods and approaches used in this dissertation will prove to be dynamic and effective in predicting WNV for future disease modelers, abatement personnel, and public health officials.

REFERENCES

Allan, B. F., R. B. Langerhans, W. A. Ryberg, W. J. Landesman, N. W. Griffin, R. S. Katz, B. J. Oberle, M. R. Schutzenhofer, K. N. Smyth, A. De St. Maurice, L. Clark, K. R. Crooks, D. E. Hernandez, R. G. McLean, R. S. Ostfeld, and J. M. Chase. **2009**. Ecological correlates of risk and incidence of West Nile virus in the United States. Oecologia. 158: 699–708.

Alto, B. W., and S. A. Juliano. **2001**. Precipitation and Temperature Effects on Populations of Aedes albopictus (Diptera: Culicidae): Implications for Range Expansion . J. Med. Entomol.

Andreadis, T. G., J. F. Anderson, and C. R. Vossbrinck. **2001**. Mosquito surveillance for West Nile virus in Connecticut, 2000: Isolation from Culex pipiens, Cx. restuans, Cx. salinarius, and Culiseta melanura. Emerg. Infect. Dis. 7: 670–674.

Andreadis, T. G., J. F. Anderson, C. R. Vossbrinck, and A. J. Main. **2004**. Epidemiology of West Nile virus in Connecticut: A five-year analysis of mosquito data 1999-2003. Vector-Borne Zoonotic Dis.

Armstrong, P. M., T. G. Andreadis, J. J. Shepard, and M. C. Thomas. **2017**. Northern range expansion of the Asian tiger mosquito (Aedes albopictus): Analysis of mosquito data from Connecticut, USA. PLoS Negl. Trop. Dis.

Babyak, M. A. 2004. What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. Psychosom. Med. 66: 411–421.

Bale, J. S., and S. A. L. Hayward. **2010**. Insect overwintering in a changing climate. J. Exp. Biol.

Bansal, S., G. Chowell, L. Simonsen, A. Vespignani, and C. Viboud. **2016**. Big data for infectious disease surveillance and modeling. J. Infect. Dis. 214: S375–S379.

Barker, C. M. 2019. Models and Surveillance Systems to Detect and Predict West Nile Virus Outbreaks. J. Med. Entomol. 56: 1508–1515.

Bell, J. A., N. J. Mickelson, and J. A. Vaughan. **2005**. West Nile virus in host-seeking mosquitoes within a residential neighborhood in Grand Forks, North Dakota. Vector-Borne Zoonotic Dis.

Bendel, R. B., and A. A. Afifi. **1977**. Comparison of stopping rules in forward "stepwise" regression. J. Am. Stat. Assoc. 72: 46–53.

Bennett, L. 2018. Deforestation and Climate Change – Climate Institute. Clim. Inst.

Betoret, E., and N. Betoret. **2020**. Chapter 10 - Globalization of technologies: Pros and cons. *In* Sustain. Food Syst.

Birch, C. P. D., S. P. Oom, and J. A. Beecham. **2007**. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecol. Modell. 206: 347–359.

Blitvich, B. J. 2008. Transmission dynamics and changing epidemiology of West Nile virus. Anim. Health Res. Rev.

Bongaarts, J. 2009. Human population growth and the demographic transition. Philos. Trans. R. Soc. B Biol. Sci.

Bonizzoni, M., G. Gasperi, X. Chen, and A. A. James. **2013**. The invasive mosquito species Aedes albopictus: Current knowledge and future perspectives. Trends Parasitol.

Bowden, S. E., K. Magori, and J. M. Drake. **2011**. Regional differences in the association between land cover and West Nile virus disease incidence in humans in the United States. Am. J. Trop. Med. Hyg. 84: 234–238.

Bowen, R. A., and N. M. Nemeth. **2007**. Experimental infections with West Nile virus. Curr. Opin. Infect. Dis.

Bursac, Z., C. H. Gauss, D. K. Williams, and D. W. Hosmer. **2008**. Purposeful selection of variables in logistic regression. Source Code Biol. Med. 3: 1–8.

Byrne, K., and R. A. Nichols. **1999**. Culex pipiens in London Underground tunnels: Differentiation between surface and subterranean populations. Heredity (Edinb).

Caglar, S. S., B. Alten, R. Bellini, F. M. Simsek, and S. Kaynas. **2003**. Comparison of nocturnal activities of mosquitoes (Diptera: Culicidae) sampled by New Jersey light traps and CO2 traps in Belek, Turkey. J. Vector Ecol.

Carson, P. J., S. M. Borchardt, B. Custer, H. E. Prince, J. Dunn-Williams, V. Winkelman, L. Tobler, B. J. Biggerstaff, R. Lanciotti, L. R. Petersen, and M. P. Busch. **2012**. Neuroinvasive disease and west nile virus infection, North Dakota, USA, 1999-2008. Emerg. Infect. Dis.

CDC. **2013**. West Nile Virus in the United States: Guidelines for Surveillance, Prevention, and Control. Appendix 2: Calculation and Application of a Vector Index (VI) Reflecting the Number of WN Virus Infected Mosquitoes in a Population. 4th Revision.

CDC. **2015**. National Notifiable Disease Surveillance System (NNDSS): Arboviral Diseases, Neuroinvasive and Non-neuroinvasive 2015 Case Definition.

CDC. **2017**. West Nile Virus - Transmission. Centers Dis. Control Prev.

CDC. **2018a**. One Health Basics. Centers Dis. Control Prev. Heal.

CDC. **2018b**. Final cumulative maps and data | West Nile Virus | CDC. Webpage.

CDC. **2018c**. West Nile virus Surveillance Resources: ArboNET.

CDC. **2018d**. WNV Symptoms, Diagnosis, & Treatment. (https://www.cdc.gov/westnile/symptoms/index.html).

CDC. **2019**. National Notifiable Diseases Surveillance Systems (NNDSS): MMWR Week Fact Sheet.

CDC. **2020a**. West Nile virus fact sheet.

CDC. **2020b**. National arboviral surveillance system (ArboNET), Arboviral Diseases Branch, Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention (CDC). **2010**. Morbidity and Mortality Weekly Report Surveillance for Human West Nile Virus Disease — United States , 1999 – 2008. MMWR Wkly. Rep. 59: 1999–2008.

Cohen, J. M., D. J. Civitello, A. J. Brace, E. M. Feichtinger, C. N. Ortega, J. C. Richardson, E. L. Sauer, X. Liu, and J. R. Rohr. **2016**. Spatial scale modulates the strength of ecological processes driving disease distributions. Proc. Natl. Acad. Sci. U. S. A.

Colpitts, T. M., M. J. Conway, R. R. Montgomery, and E. Fikrig. **2012**. West Nile virus: Biology, transmission, and human infection. Clin. Microbiol. Rev. 25: 635–648.

Condeso, T Emiko, Meentemeyer, R. 2007. Effects of Landscape Heterogeneity on the Emerging Forest Disease Sudden Oak Death. J. Ecol. 95: 364–375.

County, C. 2019. Cook County Government Open Data. Cook Cty. Data Cat.

Crutzen, P. J. 2006. The anthropocene. *In* Earth Syst. Sci. Anthr.

Day, J. F., and J. Shaman. **2008**. Using Hydrologic Conditions to Forecast the Risk of Focal and Epidemic Arboviral Transmission in Peninsular Florida. J. Med. Entomol.

DeGroote, J. P., and R. Sugumaran. **2012**. National and regional associations between human West Nile virus incidence and demographic, landscape, and land use conditions in the coterminous United States. Vector-Borne Zoonotic Dis.

DeGroote, J. P., R. Sugumaran, S. M. Brend, B. J. Tucker, and L. C. Bartholomay. **2008**. Landscape, demographic, entomological, and climatic associations with human disease incidence of West Nile virus in the state of Iowa, USA. Int. J. Health Geogr.

DeGroote, J. P., R. Sugumaran, and M. Ecker. **2014**. Landscape, demographic and climatic associations with human West Nile virus occurrence regionally in 2012 in the United States of America. Geospat. Health. 9: 153–168.

Environmental Systems Research Insititute. **2011**. ArcGIS Desktop.

Epstein, P. R., and C. Defilippo. **2001**. West Nile virus and drought. Glob. Chang. Hum. Heal. 2: 105–107.

Falchi, F., Cinzano, P., Duriscoe, D., Kyba, C. C. M., Elvidge, C. D., Baugh, K., Portnov, B., Rybnikova, N. A., Furgoni, R. 2016a. The new world atlas of artificial night sky brightness. Sci. Adv.

Falchi, F., Cinzano, P., Duriscoe, D., Kyba, C. C. M., Elvidge, C. D., Baugh, K., Portnov, B., Rybnikova, N. A., Furgoni, R. 2016b. Supplement to the New World Atlas of Artificial Night Sky Brightness. GFZ Data Serv.

FOSTER, W. A., and E. D. WALKER. **2002**. MOSQUITOES (Culicidae). *In* Med. Vet. Entomol.

Fraterrigo, J. M., A. B. Langille, and J. A. Rusak. **2020**. Stochastic disturbance regimes alter patterns of ecosystem variability and recovery. PLoS One. 15: 1–20.

Geery, P. R., and R. E. Holub. **1989**. Seasonal abundance and control of Culex spp. in catch basins in Illinois. J. Am. Mosq. Control Assoc.

Gibbs, S., Wimberly, Michael, Madden, Marguerite, Masour, J., Yabsley, Michael, Stallknecht, D. 2006. Factors Affecting the Geographic Distribution of West Nile Virus in Georgia, USA: 2002–2004. Vector-Borne Zoonotic Dis. 6: 73–82.

Giordano, B. V., S. Kaur, and F. F. Hunter. **2017**. West Nile virus in Ontario, Canada: A twelve-year analysis of human case prevalence, mosquito surveillance, and climate data. PLoS One. 12: 1–15.

Goddard, L. B., A. E. Roth, W. K. Reisen, and T. W. Scott. **2002**. Vector competence of California mosquitoes for West Nile virus. Emerg. Infect. Dis. 8: 1385–1391.

Granwehr, B. P., K. M. Lillibridge, S. Higgs, P. W. Mason, J. F. Aronson, G. A. Campbell, and A. D. T. Barrett. **2004**. West Nile virus: Where are we now? Lancet Infect. Dis. 4: 547–556.

Gray, T. J., and C. E. Webb. **2014**. A review of the epidemiological and clinical aspects of West Nile virus. Int. J. Gen. Med. 7: 193–203.

Greenland, S., and R. M. Mickey. **1989**. The impact of confounder selection criteria on effect estimation. Am. J. Epidemiol. 130: 1066.

Greer, A., V. Ng, and D. Fisman. **2008**. Climate change and infectious diseases in North America: The road ahead. CMAJ.

Griggs, D. J., and M. Noguer. **2002**. Climate change 2001: The scientific basis. Contribution of working group I to the third assessment report of the intergovernmental panel on climate change. Weather.

Gubler, D. J. 1998. Resurgent vector-borne diseases as a global health problem. *In* Emerg. Infect. Dis.

Hadfield, J., A. F. Brito, D. M. Swetnam, C. B. F. Vogels, R. E. Tokarz, K. G. Andersen, R. C. Smith, T. Bedford, and N. D. Grubaugh. **2019**. Twenty years of West Nile virus spread and evolution in the Americas visualized by Nextstrain. PLoS Pathog. 15: 1–18.

Hahn, M. B., A. J. Monaghan, M. H. Hayden, R. J. Eisen, M. J. Delorey, N. P. Lindsey, R. S. Nasci, and M. Fischer. **2015**. Meteorological conditions associated with increased incidence of west nile virus disease in the United States, 2004-2012. Am. J. Trop. Med. Hyg. 92: 1013–1022.

Haines, A., P. Scheelbeek, and K. Abbasi. **2019**. Challenges for health in the Anthropocene epoch. BMJ.

Hamer, G. L., U. D. Kitron, J. D. Brawn, S. R. Loss, M. O. Ruiz, T. L. Goldberg, and E. D. Walker. **2008**. <I>Culex pipiens</I> (Diptera: Culicidae): A Bridge Vector of West Nile Virus to Humans. J. Med. Entomol.

Hamer, G. L., U. D. Kitron, T. L. Goldberg, J. D. Brawn, S. R. Loss, M. O. Ruiz, D. B. Hayes, and E. D. Walker. **2009**. Host selection by Culex pipiens mosquitoes and west nile virus amplification. Am. J. Trop. Med. Hyg. 80: 268–278.

Harbison, J. E., M. Henry, C. Xamplas, and L. R. Dugas. **2014**. Evaluation of culex pipiens populations in a residential area with a high density of catch basins in a suburb of Chicago, Illinois. J. Am. Mosq. Control Assoc.

Hawkins, D. M. 2004. The Problem of Overfitting. J. Chem. Inf. Comput. Sci. 44: 1–12.

HAYES, C. G. 2006. West Nile Virus: Uganda, 1937, to New York City, 1999. Ann. N. Y. Acad. Sci.

Hayes, E. B., N. Komar, R. S. Nasci, S. P. Montgomery, D. R. O'Leary, and G. L. Campbell. **2005**. Epidemiology and transmission dynamics of West Nile virus disease. Emerg. Infect. Dis.

Huang, S., G. Molaei, and T. G. Andreadis. **2008**. Genetic insights into the population structure of Culex pipiens (Diptera: Culicidae) in the northeastern United States by using microsatellite analysis. Am. J. Trop. Med. Hyg.

Hurlbert, A. H., and W. Jetz. **2007**. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. Proc. Natl. Acad. Sci. U. S. A. 104: 13384–13389.

Ipcc. **2013**. Working Group I Contribution to the IPCC Fifth Assessment Report, Climate Change 2013: The Physical Science Basis. Ipcc.

IPCC. **2018**. Summary for Policymakers of IPCC Special Report on Global Warming of 1.5°C approved by governments, Ipcc.

Irwin, P., C. Arcari, J. Hausbeck, and S. Paskewitz. **2008**. Urban wet environment as mosquito habitat in the upper Midwest. Ecohealth.

Jepsen, J. U., S. B. Hagen, R. A. Ims, and N. G. Yoccoz. **2008**. Climate change and outbreaks of the geometrids Operophtera brumata and Epirrita autumnata in subarctic birch forest: Evidence of a recent outbreak range expansion. J. Anim. Ecol.

Johnson, P. T. J., R. S. Ostfeld, and F. Keesing. **2015**. Frontiers in research on biodiversity and disease. Ecol. Lett. 18: 1119–1133.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. **2008**. Global trends in emerging infectious diseases. Nature.

Juliano, S. A., and L. Philip Lounibos. **2005**. Ecology of invasive mosquitoes: Effects on resident species and on human health. Ecol. Lett.

Karki, S., W. M. Brown, J. Uelmen, M. O'Hara Ruiz, and R. L. Smith. **2020**. The drivers of West Nile virus human illness in the Chicago, Illinois, USA area: Fine scale dynamic effects of weather, mosquito infection, social, and biological conditions. PLoS One.

Keesing, F., R. D. Holt, and R. S. Ostfeld. **2006**. Effects of species diversity on disease risk. Ecol. Lett.

Keyel, A. C., O. E. Timm, P. Bryon Backenson, C. Prussing, S. Quinones, K. A. McDonough, M. Vuille, J. E. Conn, P. M. Armstrong, T. G. Andreadis, and L. D. Kramer. **2019**. Seasonal temperatures and hydrological conditions improve the prediction of West Nile virus infection rates in Culex mosquitoes and human case counts in New York and Connecticut, PLoS One.

Kilpatrick, A. M., P. Daszak, M. J. Jones, P. P. Marra, and L. D. Kramer. **2006**. Host heterogeneity dominates West Nile virus transmission. Proc. R. Soc. B Biol. Sci.

Kilpatrick, A. M., L. D. Kramer, M. J. Jones, P. P. Marra, and P. Daszak. **2006**. West Nile virus epidemics in North America are driven by shifts in mosquito feeding behavior. PLoS Biol. 4: 606–610.

Kilpatrick, A. M., S. L. LaDeau, and P. P. Marra. **2007**. Ecology of West Nile virus transmission and its impact on birds in the western hemisphere. Auk.

Kilpatrick, A. M., and W. J. Pape. **2013**. Predicting human west nile virus infections with mosquito surveillance data. Am. J. Epidemiol. 178: 829–835.

Knies, J. L., and J. G. Kingsolver. **2010**. Notes and comments erroneous arrhenius: Modified Arrhenius model best explains the temperature dependence of ectotherm fitness. Am. Nat.

Komar, N., S. Langevin, S. Hinten, N. Nemeth, E. Edwards, D. Hettler, B. Davis, R. Bowen, and M. Bunning. **2003**. Experimental infection of North American birds with the New York 1999 strain of West Nile virus. Emerg. Infect. Dis.

Kramer, L. D., and A. T. Ciota. **2015**. Dissecting vectorial capacity for mosquito-borne viruses. Curr. Opin. Virol.

Kramer, L. D., L. M. Styer, and G. D. Ebel. **2008**. A Global Perspective on the Epidemiology of West Nile Virus. Annu. Rev. Entomol. 53: 61–81.

Kuhn, K., D. Campbell-Lendrum, A. Haines, and J. Cox. **2005**. Using climate to predict infectious disease epidemics. Geneva World Heal. Organ. 54.

Lafferty, K. D. (Us G. S. 2010. CONCEPTS & SYNTHESIS: The ecology of climate change and infectious diseases. Ecology. 90: 888–900.

Lanciotti, R. S., J. T. Roehrig, V. Deubel, J. Smith, M. Parker, K. Steele, B. Crise, K. E. Volpe, M. B. Crabtree, J. H. Scherret, R. A. Hall, J. S. MacKenzie, C. B. Cropp, B. Panigrahy, E. Ostlund, B. Schmitt, M. Malkinson, C. Banet, J. Weissman, N. Komar, H. M. Savage, W. Stone, T. McNamara, and D. J. Gubler. **1999**. Origin of the West Nile virus responsible for an outbreak of encephalitis in the Northeastern United States. Science (80-.).

Lawton, J. H. 2016. Nordic Society Oikos Are There General Laws in Ecology ? Author (s): John H . Lawton Published by : Wiley on behalf of Nordic Society Oikos Stable URL : http://www.jstor.org/stable/3546712 Accessed : 16-05-2016 03 : 36 UTC Your use of the JSTOR archive. Wiley behalf Nord. Soc. Oikos. 84: 177–192.

Lei, J., M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman. **2018**. Distribution-Free Predictive Inference for Regression. J. Am. Stat. Assoc. 113: 1094–1111.

Lever, J., M. Krzywinski, and N. Altman. **2016**. Points of Significance: Model selection and overfitting. Nat. Methods. 13: 703–704.

Levin, S. A. 1992. The problem of pattern and scale in ecology. Ecology. 73: 1943–1967.

Lindsey, R. 2020. Climate Change: Atmospheric Carbon Dioxide. Nature.

Loss, S. R., G. L. Hamer, E. D. Walker, M. O. Ruiz, T. L. Goldberg, U. D. Kitron, and J. D. Brawn. **2009**. Avian host community structure and prevalence of West Nile virus in Chicago, Illinois. Oecologia. 159: 415–424.

Madder, D. J., G. A. Surgeoner, and B. V. Helson. **1983**. Number of generations, egg production, and developmental time of Culex pipiens and Culex restuans (Diptera: Culicidae) in southern Ontario. J. Med. Entomol.

Manore, C. A., J. K. Davis, R. C. Christofferson, D. M. Wesson, J. M. Hyman, and C. N. Mores. **2014**. Towards an Early Warning System for Forecasting Human West Nile Virus Incidence. PLoS Curr. 1–21.

Mason, B. J. 1989. The greenhouse effect. Contemp. Phys.

Mayer, A. L., and G. N. Cameron. **2003**. Consideration of grain and extent in landscape studies of terrestrial vertebrate ecology. Landsc. Urban Plan. 65: 201–217.

Mccall, P. J., and G. Eaton. **2001**. Olfactory memory in the mosquito Culex quinquefasciatus. Med. Vet. Entomol.

McLean, R. G., S. R. Ubico, D. Bourne, and N. Komar. **2002**. West Nile virus in livestock and wildlife. Curr. Top. Microbiol. Immunol.

McMichael, A. J., D. H. Campbell-Lendrum, C. F. Corvalán, K. L. Ebi, A. K. Githeko, J. D. Scheraga, and A. Woodward. **2003**. Climate change and infectious diseases. Climate change and human health: risks and responses, World Heal. Organ.

Meentemeyer, R. K., S. E. Haas, and T. Václavík. **2012**. Landscape Epidemiology of Emerging Infectious Diseases in Natural and Human-Altered Ecosystems. Annu. Rev. Phytopathol. 50: 379–402.

Molaei, G., T. G. Andreadis, P. M. Armstrong, J. F. Anderson, and C. R. Vossbrinck. **2006**. Host feeding patterns of Culex mosquitoes and west nile virus transmission, northeastern United States. Emerg. Infect. Dis. 12: 468–474.

Moore, C. G., and C. J. Mitchell. **1997**. Aedes albopictus in the United States: Ten-Year Presence and Public Health Implications. Emerg. Infect. Dis.

Morin, C. W., and A. C. Comrie. **2013**. Regional and seasonal response of a West Nile virus vector to climate change. Proc. Natl. Acad. Sci. U. S. A. 110: 15620–15625.

Murray, K. O., E. Mertens, and P. Desprès. **2010**. West Nile virus and its emergence in the United States of America. Vet. Res.

Myers, S. S., L. Gaffikin, C. D. Golden, R. S. Ostfeld, K. H. Redford, T. H. Ricketts, W. R. Turner, and S. A. Osofsky. **2013**. Human health impacts of ecosystem alteration. Proc. Natl. Acad. Sci. U. S. A.

Nasci, R. S., and J. P. Mutebi. **2019**. Reducing West Nile Virus Risk Through Vector Management. J. Med. Entomol. 56: 1516–1521.

Nelson, C. A., S. Saha, K. J. Kugeler, M. J. Delorey, M. B. Shankar, A. F. Hinckley, and P. S. Mead. **2015**. Incidence of clinician-diagnosed lyme disease, United States, 2005–2010. Emerg. Infect. Dis. 21: 1625–1631.

Ng, T., D. Hathaway, N. Jennings, D. Champ, Y. W. Chiang, and H. J. Chu. **2003**. Equine vaccine for West Nile virus. Dev. Biol. (Basel).

Oregon State University. **2019**. PRISM Climate Group. (http://prism.oregonstate.edu).

Ostfeld, R. S., and F. Keesing. **2000**. Biodiversity and disease risk: The case of Lyme disease. Conserv. Biol. 14: 722–728.

Patz, J. A. 1996. Global Climate Change and Emerging Infectious Diseases. JAMA J. Am. Med. Assoc.

Patz, J. A., D. Campbell-Lendrum, T. Holloway, and J. A. Foley. **2005**. Impact of regional climate change on human health. Nature. 438: 310–317.

Patz, J. A., A. K. Githeko, J. P. McCarty, S. Hussein, and U. Confalonieri. **2003**. Chapter 6: Climate change and infectious diseases. *In* Clim. Chang. Hum. Heal.

Patz, J. A., and W. K. Reisen. **2001**. Immunology, climate change and vector-borne diseases. Trends Immunol.

Petersen, L. R. 2001. West Nile Virus West Nile Virus: A Reemerging Global Pathogen. 7: 611–614.

Petersen, L. R., A. C. Brault, and R. S. Nasci. **2013**. West Nile virus: Review of the literature. JAMA - J. Am. Med. Assoc. 310: 308–315.

Petersen, L. R., P. J. Carson, B. J. Biggerstaff, B. Custer, S. M. Borchardt, and M. P. Busch. **2013**. Estimated cumulative incidence of West Nile virus infection in US adults, 1999-2010. Epidemiol. Infect. 141: 591–595.

Petersen, L. R., and M. Fischer. **2012**. Unpredictable and difficult to control - The adolescence of West Nile virus. N. Engl. J. Med.

Raval, A., and V. Ramanathan. **1989**. Observational determination of the greenhouse effect. Nature.

Read, N. R., J. R. Rooker, and J. P. Gathman. **1994**. Public perception of mosquito annoyance measured by a survey and simultaneous mosquito sampling. J. Am. Mosq. Control Assoc. 10: 79–87.

Reisen, W. K., Y. Fang, and V. M. Martinez. **2006**. Effects of Temperature on the Transmission of West Nile Virus by <I>Culex tarsalis</I> (Diptera: Culicidae). J. Med. Entomol.

Reisen, W., H. Lothrop, R. Chiles, M. Madon, C. Cossen, L. Woods, S. Husted, V. Kramer, and J. Edman. **2004**. West Nile virus in California. Emerg. Infect. Dis.

Reither, P. 2001. Climate Change and Mosquito-Borne Disease. Environ. Int. 109: 141–161.

Robinet, C., and A. Roques. **2010**. Direct impacts of recent climate warming on insect populations. Integr. Zool.

Roiz, D., S. Ruiz, R. Soriguer, and J. Figuerola. **2014**. Climatic effects on mosquito abundance in Mediterranean wetlands. Parasites and Vectors. 7: 1–13.

Rosà, R., G. Marini, L. Bolzoni, M. Neteler, M. Metz, L. Delucchi, E. A. Chadwick, L. Balbo, A. Mosca, M. Giacobini, L. Bertolotti, and A. Rizzoli. **2014**. Early warning of West Nile virus mosquito vector: Climate and land use models successfully explain phenology and abundance of Culex pipiens mosquitoes in north-western Italy. Parasites and Vectors. 7: 1–12.

Rosenberg, R., N. P. Lindsey, M. Fischer, C. J. Gregory, A. F. Hinckley, P. S. Mead, G. Paz-Bailey, S. H. Waterman, N. A. Drexler, G. J. Kersh, H. Hooks, S. K. Partridge, S. N. Visser, C. B. Beard, and L. R. Petersen. **2018**. Vital signs: Trends in reported vectorborne disease cases — United States and Territories, 2004-2016. Morb. Mortal. Wkly. Rep. 67: 496–501.

Rueda, L. M., K. J. Patel, R. C. Axtell, and R. E. Stinner. **1990**. Temperature-dependent development and survival rates of Culex quinquefasciatus and Aedes aegypti (Diptera: Culicidae). J. Med. Entomol.

Ruiz, M. O., L. F. Chaves, G. L. Hamer, T. Sun, W. M. Brown, E. D. Walker, L. Haramis, T. L. Goldberg, and U. D. Kitron. **2010**. Local impact of temperature and precipitation on West Nile virus infection in Culex species mosquitoes in northeast Illinois, USA. Parasites and Vectors.

Ruiz, M. O., C. Tedesco, T. J. McTighe, C. Austin, and K. Uriel. **2004**. Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. Int. J. Health Geogr. 3.

Ruiz, M. O., E. D. Walker, E. S. Foster, L. D. Haramis, and U. D. Kitron. **2007**. Association of West Nile virus illness and urban landscapes in Chicago and Detroit. Int. J. Health Geogr.

Russell, C., and F. F. Hunter. **2012**. Culex pipiens (Culicidae) is attracted to humans in southern Ontario, but will it serve as a bridge vector of West Nile virus? Can. Entomol.

Ryan, S. J., C. J. Carlson, E. A. Mordecai, and L. R. Johnson. **2018**. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. PLoS Negl. Trop. Dis.

Schwarz, G. 1978. Estimating the Dimension of a Model. Ann. Stat. 6: 461–464.

Sejvar, J. J. 2003. West Nile virus: An historical overview. Ochsner J. 5: 6–10.

Semenza, J. C., and B. Menne. **2009**. Climate change and infectious diseases in Europe. Lancet Infect. Dis. 9: 365–375.

Skaff, N. K., and K. S. Cheruvelil. **2016**. Fine-scale wetland features mediate vector and climate-dependent macroscale patterns in human West Nile virus incidence. Landsc. Ecol. 31: 1615–1628.

Smith, K. F., M. Goldberg, S. Rosenthal, L. Carlson, J. Chen, C. Chen, and S. Ramachandran. **2014**. Global rise in human infectious disease outbreaks. J. R. Soc. Interface.

Smithburn, K. C., T. P. Hughes, A. W. Burke, and J. H. Paul. **1940**. A Neurotropic Virus Isolated from the Blood of a Native of Uganda 1. Am. J. Trop. Med. Hyg. s1-20: 471–492.

Solomon, S., G. K. Plattner, R. Knutti, and P. Friedlingstein. **2009**. Irreversible climate change due to carbon dioxide emissions. Proc. Natl. Acad. Sci. U. S. A.

Solomon, T. 2004. Flavivirus encephalitis. N. Engl. J. Med.

Su, T., and M. S. Mulla. **2001**. Effects of temperature on development, mortality, mating and blood feeding behavior of Culiseta incidens (Diptera: Culicidae). J. Vector Ecol.

The Lancet. **2019**. Planetary health in the Anthropocene. Lancet.

Tibshirani, R., J. Lei, M. G. Sell, A. Rinaldo, J. Taylor, and R. Tibshirani. **2018**. LOCO : The Good , the Bad , and the Ugly What can we do without a model ?

Turell, M. J., D. J. Dohm, M. R. Sardelis, M. L. O'Guinn, T. G. Andreadis, and J. A. Blow. **2005**. An update on the potential of North American mosquitoes (Diptera: Culicidae) to transmit West Nile virus. J. Med. Entomol.

Uelmen, J.A., Irwin, P., Brown, W., Karki, S., Ruiz, M.O., Li, B., Smith, R. 2020. Assessing Ultra-Fine-Scale Factors to Improve Human West Nile Virus Disease Models in the Chicago Area.

Uelmen, J. A., P. Irwin, D. Bartlett, W. M. Brown, S. Karki, M. O. Ruiz, J. M. Fraterrigo, B. Li, and R. L. Smith. **2020**. Effects of Scale on Modeling West Nile Virus Disease Risk. Under Rev.

US EPA. **2016**. Climate change indicators in the United States 2016 Fourth Edition. Clim. Chang. Indic. United States 2016.

USDA. **2018**. National Agriculture Imagery Program (NAIP) Imagery. (https://catalog.data.gov/dataset/national-geospatial-data-asset-ngda-naip-imagery).

USGS. **2011**. National Geospatial Data Asset (NGDA) Land Use Land Cover.

USGS. **2019**. EarthExplorer. (https://earthexplorer.usgs.gov/).

Vezzani, D. 2007. Review: Artificial container-breeding mosquitoes and cemeteries: A perfect match. Trop. Med. Int. Heal.

Warren, R., J. Price, J. VanDerWal, S. Cornelius, and H. Sohl. **2018**. The implications of the United Nations Paris Agreement on climate change for globally significant biodiversity areas. Clim. Change.

Waterman, S. H., H. S. Margolis, and J. J. Sejvar. **2015**. Surveillance for dengue and dengue-associated neurologic syndromes in the United States. Am. J. Trop. Med. Hyg. 92: 996–998.

Whitmee, S., A. Haines, C. Beyrer, F. Boltz, A. G. Capon, B. F. De Souza Dias, A. Ezeh, H. Frumkin, P. Gong, P. Head, R. Horton, G. M. Mace, R. Marten, S. S. Myers, S. Nishtar, S. A. Osofsky, S. K. Pattanayak, M. J. Pongsiri, C. Romanelli, A. Soucat, J. Vega, and D. Yach. **2015**. Safeguarding human health in the Anthropocene epoch: Report of the Rockefeller Foundation-Lancet Commission on planetary health. Lancet.

WHO. **2017**. Vector-borne Diseases Fact Sheet. WHO.

Wiens, J. A. 1989. Spatial scaling in ecology. Funct. Ecol. 3: 385–397.

Wimberly, M. C., M. B. Hildreth, S. P. Boyte, E. Lindquist, and L. Kightlinger. **2008**. Ecological niche of the 2003 West Nile virus epidemic in the northern Great Plains of the United States. PLoS One.

Wimberly, M. C., A. Lamsal, P. Giacomo, and T. W. Chuang. **2014**. Regional variation of climatic influences on West Nile virus outbreaks in the United States. Am. J. Trop. Med. Hyg. 91: 677–684.

Yee, D. A., and S. A. Juliano. **2006**. Consequences of detritus type in an aquatic microsystem: Effects on water quality, micro-organisms and performance of the dominant consumer. Freshw. Biol.

Zheng, X., D. Zhang, Y. Li, C. Yang, Y. Wu, X. Liang, Y. Liang, X. Pan, L. Hu, Q. Sun, X. Wang, Y. Wei, J. Zhu, W. Qian, Z. Yan, A. G. Parker, J. R. L. Gilles, K. Bourtzis, J. Bouyer, M. Tang, B. Zheng, J. Yu, J. Liu, J. Zhuang, Z. Hu, M. Zhang, J. T. Gong, X. Y. Hong, Z. Zhang, L. Lin, Q. Liu, Z. Hu, Z. Wu, L. A. Baton, A. A. Hoffmann, and Z. Xi. **2019**. Incompatible and sterile insect techniques combined eliminate mosquitoes. Nature.

APPENDIX A: SUPPLEMENTARY MATERIALS FOR CHAPTER 2

Table A.1. Estimates of differences in cumulative monthly MIR and VI by each of four model scenarios: null, global, county-specific best fit from county-specific best-fit values. Top cells indicate mean values, while bottom cells indicate cumulative (2000-2018) values for each respective outcome variable. Cells labeled no estimate (denoted *N.E.*) are a result of no or too few of mosquito abundance data to capture a reliable vector index value.

Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?
preci	1	1	1.00	No	asianpct	3	2	1.50	No	dhipct	1	2	0.50	No	dospct	1	2	0.50	No
preci	1	1	1.00	No	asianpct	1	2	0.50	No	dhipct	1	2	0.50	No	dospct	1	2	0.50	No
preci	1	1	1.00	No	asianpct	3	2	1.50	No	dhipct	1	2	0.50	No	dospct	1	2	0.50	No
preci	1	1	1.00	No	asianpct	1	2	0.50	No	dhipct	1	2	0.50	No	dospct	1	2	0.50	No
preci	1	1	1.00	No	asianpct	1	2	0.50	No	dhipct	1	2	0.50	No	dospct	1	2	0.50	No
preci	1	1	1.00	No	blackpct	1	2	0.50	No	dhipct	1	2	0.50	No	dospct	1	2	0.50	No
preci	1	1	1.00	No	blackpct	1	2	0.50	No	dhipct	1	2	0.50	No	dtotpct	1	2	0.50	No
preci	1	1	1.00	No	blackpct	1	2	0.50	No	dhipct	1	2	0.50	No	dtotpct	1	2	0.50	No
preci	2	1	2.00	No	blackpct	1	2	0.50	No	dhipct	1	2	0.50	No	dtotpct	1	2	0.50	No
preci	1	1	1.00	No	blackpct	1	2	0.50	No	dhipct	1	2	0.50	No	dtotpct	1	2	0.50	No
preci	2	1	2.00	No	blackpct	1	2	0.50	No	dlipct	1	2	0.50	No	dtotpct	1	2	0.50	No
preci	1	1	1.00	No	blackpct	1	2	0.50	No	dlipct	1	2	0.50	No	ftotpct	1	2	0.50	No
preci	1	1	1.00	No	blackpct	1	2	0.50	No	dlipct	1	2	0.50	No	ftotpct	1	2	0.50	No
preci	1	1	1.00	No	blackpct	1	2	0.50	No	dlipct	1	2	0.50	No	ftotpct	1	2	0.50	No
tempc	1	1	1.00	No	blackpct	1	2	0.50	No	dlipct	1	2	0.50	No	ftotpct	1	2	0.50	No
tempc	1	1	1.00	No	blackpct	1	2	0.50	No	dlipct	1	2	0.50	No	ftotpct	1	2	0.50	No
tempc	1	1	1.00	No	blackpct	1	2	0.50	No	dlipct	1	2	0.50	No	glandpct	1	2	0.50	No
tempc	1	1	1.00	No	blackpct	1	2	0.50	No	dlipct	1	2	0.50	No	glandpct	1	2	0.50	No
tempc	1	1	1.00	No	blackpct	1	2	0.50	No	dlipct	1	2	0.50	No	glandpct	1	2	0.50	No
tempc	1	1	1.00	No	blackpct	1	2	0.50	No	dlipct	1	2	0.50	No	glandpct	1	2	0.50	No
tempc	1	1	1.00	No	blpct	1	2	0.50	No	dlipct	1	2	0.50	No	glandpct	1	2	0.50	No
tempc	2	1	2.00	No	blpct	1	2	0.50	No	dmipct	1	2	0.50	No	glandpct	1	2	0.50	No
tempc	1	1	1.00	No	blpct	1	2	0.50	No	dmipct	1	2	0.50	No	glandpct	1	2	0.50	No
tempc	2	1	2.00	No	blpct	1	2	0.50	No	dmipct	1	2	0.50	No	glandpct	1	2	0.50	No
tempc	1	1	1.00	No	blpct	1	2	0.50	No	dmipct	1	2	0.50	No	glandpct	1	2	0.50	No
tempc	1	1	1.00	No	blpct	1	2	0.50	No	dmipct	1	2	0.50	No	glandpct	1	2	0.50	No
Yr	1	1	1.00	No	blpct	1	2	0.50	No	dmipct	1	2	0.50	No	hispanicpct	1	2	0.50	No
Yr	1	1	1.00	No	blpct	3	2	1.50	No	dmipct	1	2	0.50	No	hispanicpct	1	2	0.50	No
Yr	1	1	1.00	No	blpct	3	2	1.50	No	dmipct	3	2	1.50	No	hispanicpct	1	2	0.50	No
Yr	1	1	1.00	No	blpct	3	2	1.50	No	dmipct	4	2	2.00	No	hispanicpct	1	2	0.50	No
Yr	1	1	1.00	No	blpct	3	2	1.50	No	dmipct	1	2	0.50	No	hispanicpct	1	2	0.50	No
Yr	1	1	1.00	No	blpct	3	2	1.50	No	dmipct	3	2	1.50	No	hispanicpct	1	2	0.50	No
Yr	1	1	1.00	No	ccpct	1	2	0.50	No	dmipct	1	2	0.50	No	hispanicpct	1	2	0.50	No
Yr	1	1	1.00	No	ccpct	1	2	0.50	No	dmipct	1	2	0.50	No	hispanicpct	1	2	0.50	No
Yr	1	1	1.00	No	ccpct	1	2	0.50	No	dmipct	1	2	0.50	No	Income	1	2	0.50	No
Yr	1	1	1.00	No	dfpct	1	2	0.50	No	dmipct	1	2	0.50	No	Income	1	2	0.50	No
Yr	1	1	1.00	No	dfpct	1	2	0.50	No	dmipct	1	2	0.50	No	Income	1	2	0.50	No
# bldgs	1	2	0.50	No	dfpct	1	2	0.50	No	dmipct	1	2	0.50	No	Income	1	2	0.50	No
# bldgs	1	2	0.50	No	dfpct	1	2	0.50	No	dmipct	1	2	0.50	No	Income	1	2	0.50	No
# bldgs	1	2	0.50	No	dfpct	1	2	0.50	No	dospct	1	2	0.50	No	Income	1	2	0.50	No
asianpct	1	2	0.50	No	dfpct	1	2	0.50	No	dospct	1	2	0.50	No	Income	1	2	0.50	No
asianpct	1	2	0.50	No	dfpct	1	2	0.50	No	dospct	1	2	0.50	No	Income	1	2	0.50	No
asianpct	1	2	0.50	No	dfpct	2	2	1.00	No	dospct	1	2	0.50	No	Income	1	2	0.50	No
asianpct	1	2	0.50	No	dfpct	2	2	1.00	No	dospct	1	2	0.50	No	Jantemp	1	2	0.50	No
asianpct	1	2	0.50	No	dfpct	1	2	0.50	No	dospct	1	2	0.50	No	Jantemp	1	2	0.50	No
asianpct	1	2	0.50	No	dfpct	1	2	0.50	No	dospct	1	2	0.50	No	Jantemp	1	2	0.50	No
asianpct	1	2	0.50	No	dhipct	1	2	0.50	No	dospct	1	2	0.50	No	Jantemp	1	2	0.50	No

Table A.1. (continued)

Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?
Jantemp	1	2	0.50	No	precilag1	1	2	0.50	No	templag1	1	2	0.50	No	templag4	1	2	0.50	No	wwpct	1	2	0.50	No
Jantemp	1	2	0.50	No	precilag2	1	2	0.50	No	templag1	1	2	0.50	No	templag4	2	2	1.00	No	wwpct	1	2	0.50	No
Jantemp	1	2	0.50	No	precilag2	1	2	0.50	No	templag1	1	2	0.50	No	templag4	1	2	0.50	No	wwpct	1	2	0.50	No
Jantemp	1	2	0.50	No	precilag2	1	2	0.50	No	templag1	1	2	0.50	No	templag4	2	2	1.00	No	wwpct	1	2	0.50	No
Jantemp	1	2	0.50	No	precilag2	1	2	0.50	No	templag1	1	2	0.50	No	templag4	2	2	1.00	No	wwpct	1	2	0.50	No
Jantemp	1	2	0.50	No	precilag2	1	2	0.50	No	templag1	1	2	0.50	No	templag4	1	2	0.50	No	abund	1	3	0.33	No
Jantemp	1	2	0.50	No	precilag2	1	2	0.50	No	templag1	1	2	0.50	No	templag4	1	2	0.50	No	abund	1	3	0.33	No
mfpct	1	2	0.50	No	precilag2	1	2	0.50	No	templag1	1	2	0.50	No	templag4	2	2	1.00	No	abund	1	3	0.33	No
mfpct	1	2	0.50	No	precilag2	1	2	0.50	No	templag1	1	2	0.50	No	templag4	1	2	0.50	No	abund	1	3	0.33	No
mfpct	1	2	0.50	No	precilag2	1	2	0.50	No	templag2	1	2	0.50	No	templag4	1	2	0.50	No	abund	1	3	0.33	No
mfpct	1	2	0.50	No	precilag2	1	2	0.50	No	templag2	1	2	0.50	No	templag4	1	2	0.50	No	abund	1	3	0.33	No
mfpct	1	2	0.50	No	precilag3	1	2	0.50	No	templag2	1	2	0.50	No	templag4	1	2	0.50	No	abund	1	3	0.33	No
mfpct	1	2	0.50	No	precilag3	1	2	0.50	No	templag2	1	2	0.50	No	templag4	3	2	1.50	No	abund	1	3	0.33	No
mfpct	1	2	0.50	No	precilag3	1	2	0.50	No	templag2	1	2	0.50	No	totpop	3	2	1.50	No	abund	1	3	0.33	No
mfpct	1	2	0.50	No	precilag3	1	2	0.50	No	templag2	1	2	0.50	No	totpop	2	2	1.00	No	abund	1	3	0.33	No
mfpct	1	2	0.50	No	precilag3	1	2	0.50	No	templag2	1	2	0.50	No	totpop	1	2	0.50	No	abund	1	3	0.33	No
mfpct	1	2	0.50	No	precilag3	1	2	0.50	No	templag2	1	2	0.50	No	totpop	2	2	1.00	No	abund	1	3	0.33	No
owpct	1	2	0.50	No	precilag3	1	2	0.50	No	templag2	1	2	0.50	No	totpop	1	2	0.50	No	abund	1	3	0.33	No
owpct	1	2	0.50	No	precilag3	1	2	0.50	No	templag2	1	2	0.50	No	totpop	2	2	1.00	No	abund	1	3	0.33	No
owpct	1	2	0.50	No	precilag3	1	2	0.50	No	templag2	1	2	0.50	No	totpop	1	2	0.50	No	abund	1	3	0.33	No
owpct	1	2	0.50	No	precilag3	1	2	0.50	No	templag2	1	2	0.50	No	totpop	1	2	0.50	No	abundlag1	1	3	0.33	No
owpct	1	2	0.50	No	precilag4	2	2	1.00	No	templag2	1	2	0.50	No	totpop	1	2	0.50	No	abundlag1	1	3	0.33	No
owpct	1	2	0.50	No	precilag4	2	2	1.00	No	templag2	1	2	0.50	No	totpop	1	2	0.50	No	abundlag1	1	3	0.33	No
owpct	1	2	0.50	No	precilag4	2	2	1.00	No	templag2	1	2	0.50	No	totpop	1	2	0.50	No	abundlag1	1	3	0.33	No
owpct	2	2	1.00	No	precilag4	1	2	0.50	No	templag2	1	2	0.50	No	totpop	1	2	0.50	No	abundlag1	1	3	0.33	No
owpct	2	2	1.00	No	precilag4	2	2	1.00	No	templag2	1	2	0.50	No	totpop	1	2	0.50	No	abundlag1	1	3	0.33	No
owpct	1	2	0.50	No	precilag4	1	2	0.50	No	templag2	1	2	0.50	No	totpop	1	2	0.50	No	abundlag1	2	3	0.67	No
owpct	1	2	0.50	No	precilag4	1	2	0.50	No	templag2	1	2	0.50	No	totpop	1	2	0.50	No	abundlag1	1	3	0.33	No
pasturepct	1	2	0.50	No	precilag4	1	2	0.50	No	templag3	2	2	1.00	No	totpop	2	2	1.00	No	abundlag1	1	3	0.33	No
pasturepct	1	2	0.50	No	precilag4	1	2	0.50	No	templag3	3	2	1.50	No	totpop	1	2	0.50	No	abundlag1	1	3	0.33	No
pasturepct	1	2	0.50	No	precilag4	1	2	0.50	No	templag3	2	2	1.00	No	totpop	1	2	0.50	No	abundlag1	1	3	0.33	No
pasturepct	1	2	0.50	No	precilag4	1	2	0.50	No	templag3	2	2	1.00	No	totpop	1	2	0.50	No	abundlag1	1	3	0.33	No
pasturepct	1	2	0.50	No	precilag4	1	2	0.50	No	templag3	2	2	1.00	No	totpop	2	2	1.00	No	abundlag1	1	3	0.33	No
pasturepct	1	2	0.50	No	precilag4	2	2	1.00	No	templag3	1	2	0.50	No	whitepct	1	2	0.50	No	abundlag1	1	3	0.33	No
pasturepct	2	2	1.00	No	precilag4	1	2	0.50	No	templag3	3	2	1.50	No	whitepct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	precilag4	2	2	1.00	No	templag3	1	2	0.50	No	whitepct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	precilag4	1	2	0.50	No	templag3	3	2	1.50	No	whitepct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	precilag4	2	2	1.00	No	templag3	3	2	1.50	No	whitepct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	precilag4	1	2	0.50	No	templag3	1	2	0.50	No	whitepct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	precilag4	1	2	0.50	No	templag3	3	2	1.50	No	whitepct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	precilag4	1	2	0.50	No	templag3	1	2	0.50	No	whitepct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	shrubpct	1	2	0.50	No	templag3	3	2	1.50	No	whitepct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	shrubpct	1	2	0.50	No	templag3	2	2	1.00	No	whitepct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	1	2	0.50	No	shrubpct	1	2	0.50	No	templag3	2	2	1.00	No	whitepct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	shrubpct	1	2	0.50	No	templag3	2	2	1.00	No	whitepct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	shrubpct	1	2	0.50	No	templag3	2	2	1.00	No	wtotpct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	shrubpct	1	2	0.50	No	templag3	1	2	0.50	No	wtotpct	1	2	0.50	No	abundlag2	1	3	0.33	No
precilag1	2	2	1.00	No	shrubpct	1	2	0.50	No	templag3	1	2	0.50	No	wtotpct	1	2	0.50	No	abundlag3	1	3	0.33	No
precilag1	3	2	1.50	No	templag1	2	2	1.00	No	templag3	2	2	1.00	No	wtotpct	1	2	0.50	No	abundlag3	1	3	0.33	No
precilag1	1	2	0.50	No	templag1	1	2	0.50	No	templag3	1	2	0.50	No	wtotpct	1	2	0.50	No	abundlag3	1	3	0.33	No
precilag1	2	2	1.00	No	templag1	1	2	0.50	No	templag3	1	2	0.50	No	wwpct	1	2	0.50	No	abundlag3	1	3	0.33	No
precilag1	1	2	0.50	No	templag1	1	2	0.50	No	templag3	1	2	0.50	No	wwpct	1	2	0.50	No	abundlag3	1	3	0.33	No
precilag1	3	2	1.50	No	templag1	2	2	1.00	No	templag3	1	2	0.50	No	wwpct	1	2	0.50	No	abundlag3	1	3	0.33	No
precilag1	1	2	0.50	No	templag1	1	2	0.50	No	templag3	2	2	1.00	No	wwpct	1	2	0.50	No	abundlag3	1	3	0.33	No
precilag1	1	2	0.50	No	templag1	1	2	0.50	No	templag4	2	2	1.00	No	wwpct	1	2	0.50	No	abundlag3	1	3	0.33	No

Table A.1. (continued)

Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Observation Covariate?
abundlag3	1	3	0.33	No	hpctpostww	1	3	0.33	No	mirlag2	1	3	0.33	No	mirlag4	2	3	0.67	No
abundlag3	1	3	0.33	No	hpctpostww	1	3	0.33	No	mırlag2	1	3	0.33	No	mırlag4	2	3	0.67	No
abundlag3	1	3	0.33	No	hpctpostww	1	3	0.33	No	mirlag2	1	3	0.33	No	mirlag4	1	3	0.33	No
abundlag3	1	3	0.33	No	hpctpreww	1	3	0.33	No	mirlag2	1	3	0.33	No	mirlag4	2	3	0.67	No
abundlag3	1	3	0.33	No	hpctpreww	1	3	0.33	No	mirlag2	1	3	0.33	No	mirlag4	1	3	0.33	No
abundlag4	1	3	0.33	No	hpctpreww	1	3	0.33	No	mirlag2	1	3	0.33	No	mirlag4	2	3	0.67	No
abundlag4	1	3	0.33	No	hpctpreww	1	3	0.33	No	mırlag2	1	3	0.33	No	mırlag4	2	3	0.67	No
abundlag4	1	3	0.33	No	hpctpreww	1	3	0.33	No	mirlag2	1	3	0.33	No	mirlag4	2	3	0.67	No
abundlag4	1	3	0.33	No	hpctpreww	1	3	0.33	No	mirlag2	1	3	0.33	No	mirlag4	1	3	0.33	No
abundlag4	1	3	0.33	No	hpctpreww	1	3	0.33	No	mırlag2	1	3	0.33	No	MIRmean	1	3	0.33	No
abundlag4	2	3	0.67	No	hpctpreww	1	3	0.33	No	mirlag2	1	3	0.33	No	MIRmean	1	3	0.33	No
abundlag4	2	3	0.67	No	hpctpreww	1	3	0.33	No	mirlag2	1	3	0.33	No	MIRmean	1	3	0.33	No
abundlag4	2	3	0.67	No	hpctpreww	1	3	0.33	No	mirlag2	1	3	0.33	No	MIRmean	1	3	0.33	No
abundlag4	1	3	0.33	No	hpctpreww	1	3	0.33	No	mırlag2	1	3	0.33	No	MIRmean	1	3	0.33	No
abundlag4	2	3	0.67	No	Light pol	1	3	0.33	No	mırlag2	1	3	0.33	No	MIRmean	2	3	0.67	No
abundlag4	1	3	0.33	No	Light pol	1	3	0.33	No	mirlag2	1	3	0.33	No	MIRmean	1	3	0.33	No
abundlag4	1	3	0.33	No	Light pol	1	3	0.33	No	mirlag3	2	3	0.67	No	MIRmean	1	3	0.33	No
abundlag4	1	3	0.33	No	Light pol	1	3	0.33	No	mırlag3	1	3	0.33	No	MIRmean	1	3	0.33	No
CB	1	3	0.33	No	Light pol	1	3	0.33	No	mırlag3	2	3	0.67	No	MIRmean	1	3	0.33	No
CB	1	3	0.33	No	Light pol	1	3	0.33	No	mirlag3	2	3	0.67	No	MIRmean	1	3	0.33	No
CB	1	3	0.33	No	MIRdiff	1	3	0.33	No	mirlag3	2	3	0.67	No	MIRmean	1	3	0.33	No
CB	1	3	0.33	No	MIRdiff	1	3	0.33	No	mırlag3	1	3	0.33	No	MIRmean	1	3	0.33	No
CB	1	3	0.33	No	MIRdiff	1	3	0.33	No	mırlag3	1	3	0.33	No	resı lot area avg.	1	3	0.33	No
CB	1	3	0.33	No	MIRdiff	1	3	0.33	No	mirlag3	3	3	1.00	No	resi lot area avg.	1	3	0.33	No
hpct7089	1	3	0.33	No	MIRdiff	1	3	0.33	No	mirlag3	3	3	1.00	No	resi lot area avg.	1	3	0.33	No
hpct7089	1	3	0.33	No	MIRdiff	1	3	0.33	No	mırlag3	2	3	0.67	No	resı lot area avg.	1	3	0.33	No
hpct7089	1	3	0.33	No	MIRdiff	1	3	0.33	No	mırlag3	2	3	0.67	No	resı lot area total	1	3	0.33	No
hpct7089	1	3	0.33	No	MIRdiff	1	3	0.33	No	mirlag3	2	3	0.67	No	resi lot area total	1	3	0.33	No
hpct7089	1	3	0.33	No	MIRdiff	1	3	0.33	No	mirlag3	2	3	0.67	No	resi lot area total	1	3	0.33	No
hpct7089	1	3	0.33	No	MIRdiff	1	3	0.33	No	mırlag3	3	3	1.00	No	resı lot area total	1	3	0.33	No
hpct7089	1	3	0.33	No	MIRdiff	1	3	0.33	No	mırlag3	3	3	1.00	No	resı lot area total	1	3	0.33	No
hpct7089	1	3	0.33	No	MIRdiff	1	3	0.33	No	mirlag3	3	3	1.00	No	resi lot peri avg.	1	3	0.33	No
hpct7089	1	3	0.33	No	MIRdiff	1	3	0.33	No	mirlag3	1	3	0.33	No	resi lot peri avg.	1	3	0.33	No
hpct7089	1	3	0.33	No	mirlag1	1	3	0.33	No	mirlag3	1	3	0.33	No	resi lot peri avg.	1	3	0.33	No
hpct7089	1	3	0.33	No	mırlag1	1	3	0.33	No	mırlag3	1	3	0.33	No	resı lot peri avg.	1	3	0.33	No
hpct7089	1	3	0.33	No	mirlag1	1	3	0.33	No	mirlag3	3	3	1.00	No	resi lot peri avg.	1	3	0.33	No
hpct7089	1	3	0.33	No	mirlag1	1	3	0.33	No	mirlag3	2	3	0.67	No	resi lot peri avg.	1	3	0.33	No
hpct7089	1	3	0.33	No	mırlag1	1	3	0.33	No	mırlag3	1	3	0.33	No	resı lot peri avg.	1	3	0.33	No
hpctpost90	1	3	0.33	No	mırlag1	1	3	0.33	No	mırlag3	3	3	1.00	No	resı lot peri total	1	3	0.33	No
hpctpost90	2	3	0.67	No	mirlag1	1	3	0.33	No	mirlag3	1	3	0.33	No	resi lot peri total	1	3	0.33	No
hpctpost90	1	3	0.33	No	mirlag1	1	3	0.33	No	mirlag4	1	3	0.33	No	resi lot peri total	1	3	0.33	No
hpctpost90	1	3	0.33	No	mırlag1	1	3	0.33	No	mırlag4	3	3	1.00	No	resı lot peri total	1	3	0.33	No
hpctpost90	1	3	0.33	No	mırlag1	1	3	0.33	No	mırlag4	1	3	0.33	No	resı lot peri total	1	3	0.33	No
hpctpost90	1	3	0.33	No	mirlag1	1	3	0.33	No	mirlag4	2	3	0.67	No	VI	1	3	0.33	No
hpctpost90	1	3	0.33	No	mirlag1	1	3	0.33	No	mirlag4	1	3	0.33	No	VI	1	3	0.33	No
hpctpost90	1	3	0.33	No	mırlag1	1	3	0.33	No	mırlag4	3	3	1.00	No	VI	1	3	0.33	No
hpctpost90	1	3	0.33	No	mirlag1	1	3	0.33	No	mırlag4	3	3	1.00	No	VI	1	3	0.33	No
hpctpost90	1	3	0.33	No	mirlag1	1	3	0.33	No	mirlag4	3	3	1.00	No	VI	1	3	0.33	No
hpctpost90	1	3	0.33	No	mirlag1	1	3	0.33	No	mirlag4	3	3	1.00	No	VI	1	3	0.33	No
hpctpostww	1	3	0.33	No	mırlag1	1	3	0.33	No	mırlag4	3	3	1.00	No	VI	1	3	0.33	No
hpctpostww	1	3	0.33	No	mirlag1	1	3	0.33	No	mirlag4	1	3	0.33	No	VI	1	3	0.33	No
hpctpostww	1	3	0.33	No	mirlag1	1	3	0.33	No	mirlag4	1	3	0.33	No	VI	1	3	0.33	No
hpctpostww	1	3	0.33	No	mirlag2	1	3	0.33	No	mirlag4	1	3	0.33	No	VI	1	3	0.33	No
hpctpostww	1	3	0.33	No	mirlag2	1	3	0.33	No	mirlag4	2	3	0.67	No	VI	1	3	0.33	No
hpctpostww	1	3	0.33	No	mırlag2	1	3	0.33	No	mırlag4	2	3	0.67	No	VI	1	3	0.33	No

Table A.1. (continued)

Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Obervation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Obervation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Obervation Covariate?	Covariate	Significance Level	Data Availability/Work Load to Acquire (Score of 1,2,3,4)	Mean Covariate Value	HLC/Human Obervation Covariate?
VIlag1	1	3	0.33	No	VIlag4	1	3	0.33	No	bldg footprint area avg.	1	4	0.25	No	female obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	VIlag4	4	3	1.33	No	bldg footprint area avg.	1	4	0.25	No	female obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	VIlag4	4	3	1.33	No	bldg footprint area avg.	1	4	0.25	No	female obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	Vilag4	4	3	1.33	No	bldg footprint area avg.	1	4	0.25	No	female obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	VIlag4	4	3	1.33	No	bldg footprint area avg.	4	4	1.00	No	male obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	VIlag4	2	3	0.67	No	bldg footprint area total	1	4	0.25	No	male obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	VIlag4	1	3	0.33	No	bldg footprint area total	1	4	0.25	No	male obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	VIlag4	4	3	1.33	No	bldg footprint area total	1	4	0.25	No	male obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	VIlag4	4	3	1.33	No	bldg footprint area total	1	4	0.25	No	male obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	VIlag4	4	3	1.33	No	bldg footprint area total	1	4	0.25	No	male obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	VIlag4	1	3	0.33	No	bldg footprint area total	1	4	0.25	No	male obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	VIlag4	1	3	0.33	No	bldg footprint area total	1	4	0.25	No	male obs per visit	1	4	0.25	Yes
VIlag1	1	3	0.33	No	VIlag4	1	3	0.33	No	bldg footprint peri avg.	1	4	0.25	No	male obs per visit	1	4	0.25	Yes
VIlag2	1	3	0.33	No	VIlag4	1	3	0.33	No	bldg footprint peri avg.	1	4	0.25	No	mosquitoes per visit	1	4	0.25	Yes
VIlag2	1	3	0.33	No	VIlag4	1	3	0.33	No	bldg footprint peri avg.	1	4	0.25	No	mosquitoes per visit	1	4	0.25	Yes
VIlag2	1	3	0.33	No	VIlag4	4	3	1.33	No	bldg footprint peri avg.	1	4	0.25	No	mosquitoes per visit	1	4	0.25	Yes
VIlag2	1	3	0.33	No	VIlag4	4	3	1.33	No	bldg footprint peri total	1	4	0.25	No	mosquitoes per visit	1	4	0.25	Yes
VIlag2	1	3	0.33	No	adult obs per visit	1	4	0.25	Yes	bldg footprint peri total	1	4	0.25	No	mosquitoes per visit	1	4	0.25	Yes
VIlag2	1	3	0.33	No	adult obs per visit	1	4	0.25	Yes	bldg footprint peri total	1	4	0.25	No	mosquitoes per visit	1	4	0.25	Yes
VIlag2	1	3	0.33	No	adult obs per visit	1	4	0.25	Yes	child obs per visit	1	4	0.25	Yes	mosquitoes per visit	1	4	0.25	Yes
VIlag2	1	3	0.33	No	adult obs per visit	1	4	0.25	Yes	child obs per visit	1	4	0.25	Yes	mosquitoes per visit	1	4	0.25	Yes
VIlag2	1	3	0.33	No	adult obs per visit	1	4	0.25	Yes	child obs per visit	1	4	0.25	Yes	mosquitoes per visit	1	4	0.25	Yes
VIlag2	1	3	0.33	No	adult obs per visit	1	4	0.25	Yes	child obs per visit	1	4	0.25	Yes	mosquitoes per visit	1	4	0.25	Yes
VIlag2	1	3	0.33	No	adult obs per visit	1	4	0.25	Yes	child obs per visit	1	4	0.25	Yes	NDVI	1	4	0.25	No
VIlag2	1	3	0.33	No	adult obs per visit	1	4	0.25	Yes	child obs per visit	1	4	0.25	Yes	NDVI	1	4	0.25	No
Vilag2	1	3	0.33	No	adult obs per visit	1	4	0.25	Yes	child obs per visit	1	4	0.25	Yes	NDVI	1	4	0.25	No
VIlag3	2	3	0.67	No	avg. bldg area: avg. lot area	1	4	0.25	No	child obs per visit	1	4	0.25	Yes	NDVI	1	4	0.25	No
VIlag3	2	3	0.67	No	avg. bldg area: avg. lot area	4	4	1.00	No	child obs per visit	1	4	0.25	Yes	NDVI	1	4	0.25	No
VIlag3	2	3	0.67	No	avg. bldg area: avg. lot area	3	4	0.75	No	Culex per visit	1	4	0.25	Yes	senior obs per visit	1	4	0.25	Yes
VIlag3	2	3	0.67	No	avg. bldg area: avg. lot area	3	4	0.75	No	Culex per visit	1	4	0.25	Yes	senior obs per visit	1	4	0.25	Yes
VIlag3	2	3	0.67	No	avg. bldg area: avg. lot area	4	4	1.00	No	Culex per visit	1	4	0.25	Yes	senior obs per visit	1	4	0.25	Yes
VIlag3	2	3	0.67	No	avg. bldg area: avg. lot area	1	4	0.25	No	Culex per visit	1	4	0.25	Yes	senior obs per visit	1	4	0.25	Yes
VIlag3	2	3	0.67	No	avg. bldg peri: avg. lot area	1	4	0.25	No	Culex per visit	1	4	0.25	Yes	senior obs per visit	1	4	0.25	Yes
VIlag3	1	3	0.33	No	bldg footprint area avg.	4	4	1.00	No	Culex per visit	1	4	0.25	Yes	senior obs per visit	1	4	0.25	Yes
VIlag3	1	3	0.33	No	bldg footprint area avg.	4	4	1.00	No	Culex per visit	1	4	0.25	Yes	senior obs per visit	1	4	0.25	Yes
VIlag3	1	3	0.33	No	bldg footprint area avg.	4	4	1.00	No	Culex per visit	1	4	0.25	Yes	senior obs per visit	1	4	0.25	Yes
VIlag3	1	3	0.33	No	bldg footprint area avg.	4	4	1.00	No	Culex per visit	1	4	0.25	Yes	total bldg area: total lot area	1	4	0.25	No
VIlag3	1	3	0.33	No	bldg footprint area avg.	4	4	1.00	No	Culex per visit	1	4	0.25	Yes	total bldg area: total lot area	1	4	0.25	No
VIlag3	2	3	0.67	No	bldg footprint area avg.	4	4	1.00	No	Culex per visit	1	4	0.25	Yes	total bldg area: total lot area	1	4	0.25	No
VIlag4	1	3	0.33	No	bldg footprint area avg.	4	4	1.00	No	female obs per visit	1	4	0.25	Yes	total bldg area: total lot area	1	4	0.25	No
VIlag4	4	3	1.33	No	bldg footprint area avg.	4	4	1.00	No	female obs per visit	1	4	0.25	Yes	total bldg area: total lot area	1	4	0.25	No
															total bldg peri: total lot area	1	4	0.25	No
															total obs per visit	1	4	0.25	Yes

APPENDIX B: SUPPLEMENTARY MATERIALS FOR CHAPTER 3

Table B.1. Correlation of variation for each independent variable included in a final model. Final models, denoted by scale as “C”, “L”, and “U”, denote “County”, “Local”, and “UFS”, respectively.

Independent Variable	*Scale*	*yr*	*templag2*	*templag3*	*templag4*	*Jantemp*	*mirlag1*	*mirlag2*	*mirlag3*	*mirlag4*	*totpop*	*owpct*	*dlipct*	*dfpct*	*glandpct*	*hpctpost90*	*asianpct*	*dhipct*	*mfpct*	*wwpct*	*precilag1*	*precilag2*	*precilag4*	*Vllag1*
yr	**C**	1																						
templag2	**C**	0.0024	1																					
	C	-0.0017	0.7701	1																				
templag3	**L**			1																				
	U			1																				
	C	-0.0014	0.6402	0.7416	1																			
templag4	**L**			0.8444	1																			
	U			0.8421	1																			
	C	-0.3932	0.0933	0.1038	0.1082	1																		
Jantemp	**L**			0.0785	0.0779	1																		
mirlag1	**C**	-0.0153	0.1668	0.2053	0.2281	0.0962	1																	
mirlag2	**C**	-0.0124	0.1333	0.1784	0.2126	0.1005	0.1707	1																
mirlag3	**C**	-0.0271	0.0815	0.1292	0.1857	0.1087	0.1433	0.1753	1															
	C	-0.0314	0.0417	0.091	0.151	0.1144	0.1359	0.1525	0.1747	1														
mirlag4	**L**			0.2739	0.3131	0.2039				1														
	C	0	0.0351	0.0352	0.0337	0.0953	0.0237	0.0295	0.0272	0.0265	1													
totpop	**L**			0.0194	0.017	0.0564				0.0417	1													
owpct	**C**	-0.0006	-0.0047	-0.0057	-0.0061	0.0119	-0.0047	-0.0048	-0.0043	-0.0041	-0.1251	1												
	C	-0.0028	-0.0141	-0.0145	-0.014	-0.0347	0.0001	-0.0028	-0.0015	-0.0013	-0.1042	-0.1927	1											
dlipct	**L**			0.0096	0.0082	0.0312				0.025	0.4853		1											
dfpct	**C**	0.0039	-0.0211	-0.0218	-0.0213	-0.0478	-0.019	-0.0209	-0.0191	-0.0186	-0.2757	0.0528	-0.3254	1										
glandpct	**C**	-0.0114	-0.0179	-0.0184	-0.018	-0.0379	-0.0144	-0.0162	-0.0144	-0.0141	-0.188	0.0498	-0.2271	0.2659	1									
hpctpost90	**C**	0	-0.0308	-0.0314	-0.0305	-0.0736	-0.0309	-0.0339	-0.0309	-0.0305	-0.3075	0.0305	-0.1531	0.2036	0.2844	1								
asianpct	**U**																1							
dhipct	**U**																0.1411	1						
mfpct	**U**																-0.1988	-0.2161	1					
wwpct	**U**																-0.2279	-0.2371	0.8923	1				
precilag1	**U**																0	-0.0025	0.0048	0.0039	1			
precilag2	**U**																-0.0004	-0.0036	0.0051	0.0038	0.0496	1		
precilag4	**U**																0.0019	-0.001	0.0041	0.005	-0.1068	-0.0383	1	
Vllag1	**U**																-0.0133	0.0033	0.0076	0.0187	-0.0481	-0.0388	-0.0585	1

APPENDIX C: SUPPLEMENTARY MATERIALS FOR CHAPTER 4

Table C.1. Estimates of differences in cumulative monthly MIR and VI by each of four model scenarios: null, global, county-specific best fit from county-specific best-fit values. Top cells indicate mean values, while bottom cells indicate cumulative (2000-2018) values for each respective outcome variable. Cells labeled no estimate (denoted *N.E.*) are a result of no or too few of mosquito abundance data to capture a reliable vector index value.

County	n	Null Model				Global Model				County-Specific Global Model			
		MIR		*VI*		*MIR*		*VI*		*MIR*		*VI*	
		Difference	*%*	*Difference*	*%*	*Difference*	*%*	*Difference*	*%*	*Difference*	*%*	*Difference*	*%*
Allen, IN	4739	0.29	0.10	0.90	0.09	-1.00	-0.36	0.00	0.00	0.00	0.00	0.00	0.00
		-1192.45	-0.11	-5078.31	-0.13	-3807.12	-0.36	0.00	0.00	0.00	0.00	0.00	0.00
Anoka, MN	4996	0.09	0.01	-0.31	-0.05	2.46	0.21	0.17	0.03	0.06	0.00	0.00	0.00
		-58001.48	-4955.95	-30702.39	-5262.12	2.46	0.21	0.17	0.03	0.06	0.00	0.00	0.00
Bay, MI	2889	0.14	0.07	*N.E.*		*N.E.*		*N.E.*		0.00	0.00	*N.E.*	
		0.00	0.00	*N.E.*		*N.E.*		*N.E.*		0.00	0.00	*N.E.*	
Carroll, IA	574	-0.14	-0.56	0.00	0.23	-1.57	-6.10	0.64	49.47	0.06	0.22	*N.E.*	
		-229.87	-0.56	-5.71	-439.19	-1.57	-6.10	0.64	49.47	0.06	0.22	*N.E.*	
Carver, MN	4167	-0.44	-0.07	1.08	0.25	-1.09	-0.18	0.00	0.00	-0.28	-0.05	0.00	0.00
		-7374.06	-0.39	-10617.36	-4.16	-4012.21	-0.21	0.00	0.00	-1348.38	-0.07	0.00	0.00
Champaign, IL	158	0.08	0.09	*N.E.*		*N.E.*		*N.E.*		0.00	0.00	*N.E.*	
		0.00	0.00	*N.E.*		*N.E.*		*N.E.*		0.00	0.00	*N.E.*	
Cook, IL	1583728	-2.33	-1.06	-3.92	-0.05	-5.92	-2.68	0.01	0.00	-5.72	-2.59	0.01	0.00
		-7161135.83	-204.03	-129918882.36	-102.96	-1135726.94	-32.36	-8.25	0.00	-1106300.82	-31.52	-8.25	0.00
Dakota, MN	9194	0.00	0.00	1.75	0.20	0.98	0.13	0.00	0.00	-0.02	0.00	0.00	0.00
		-23357.21	-0.50	-53558.29	-4.62	-1876.14	-0.04	0.00	0.00	-9161.70	-0.20	0.00	0.00
Dane, WI	1679	0.07	0.25	0.02	0.27	0.98	3.58	0.00	0.00	0.03	0.12	0.00	0.00
		-39.32	-0.13	-44.55	-1.18	753.02	2.46	0.00	0.00	153.40	0.50	0.00	0.00
Dubuque, IA	2216	0.79	0.82	0.00	0.76	-0.58	-0.60	0.00	0.00	0.82	0.85	*N.E.*	
		348.25	0.48	-2.62	-6.48	411.33	0.56	0.00	0.00	700.07	0.96	*N.E.*	
DuPage, IL	814	-0.29	-0.06	*N.E.*		*N.E.*		*N.E.*		-0.02	-0.01	*N.E.*	
		-314.51	-0.09	*N.E.*		*N.E.*		*N.E.*		-19.08	-0.01	*N.E.*	
Hamilton, IN	4035	0.22	0.12	1.78	0.12	-1.47	-0.81	-0.81	-0.05	0.03	0.02	*N.E.*	
		-2882.54	-0.81	-4403.11	-0.09	-676.45	-0.19	17727.12	0.37	1249.67	0.35	*N.E.*	
Hennepin, MN	12032	-0.11	-0.01	-0.12	-0.01	1.59	0.16	0.00	0.00	-0.05	-0.01	0.00	0.00
		-39030.53	-0.47	-140304.76	-5.07	3924.15	0.05	0.00	0.00	-11486.60	-0.14	0.00	0.00
Huron, MI	114	0.07	0.02	*N.E.*		*N.E.*		*N.E.*		2.77	0.78	*N.E.*	
		7.68	0.02	*N.E.*		*N.E.*		*N.E.*		357.74	0.88	*N.E.*	
Ingham, MI	564	1.03	0.17	*N.E.*		*N.E.*		*N.E.*		0.00	0.00	*N.E.*	
		-9.85	0.00	*N.E.*		*N.E.*		*N.E.*		0.00	0.00	*N.E.*	

Table C.1. (continued)

County	n	Null Model				Global Model				County-Specific Global Model			
		MIR		*VI*		*MIR*		*VI*		*MIR*		*VI*	
		Difference	*%*	*Difference*	*%*	*Difference*	*%*	*Difference*	*%*	*Difference*	*%*	*Difference*	*%*
Lucas, OH	3376	0.14	0.01	-80.14	-0.91	4.89	0.29	0.00	0.00	-0.96	-0.06	0.00	0.00
		-16512.28	-0.42	-371062.83	-1.89	13439.81	0.34	9.68	0.00	560.13	0.01	-0.09	0.00
Macon, IL	15043	-0.04	-0.02	*N.E.*		-0.07	-0.04	*N.E.*		1.75	1.04	*N.E.*	
		-592.51	-0.02	*N.E.*		23710.54	0.94	*N.E.*		25246.41	1.00	*N.E.*	
Marion, IN	2976	0.42	0.10	0.97	0.25	0.57	0.14	0.00	0.00	1.11	0.28	0.00	0.00
		-2407.79	-0.29	-343.81	-0.04	7999.05	0.96	0.00	0.00	8057.36	0.96	0.00	0.00
Midland, MI	148	1.10	0.21	*N.E.*		*N.E.*		*N.E.*		-0.08	-0.02	*N.E.*	
		-2.35	0.00	*N.E.*		*N.E.*		*N.E.*		-9.58	-0.02	*N.E.*	
Milwaukee, WI	2744	0.20	0.20	0.80	0.17	-4.62	-4.59	-0.88	-0.19	0.02	0.02	*N.E.*	
		-108.90	-0.05	-862.12	-0.09	-2744.83	-1.31	2889.71	0.30	1246.38	0.60	*N.E.*	
Monona, IA	434	0.00	0.00	-0.31	1.73	-2.08	-0.80	-1.31	7.33	1.79	0.69	*N.E.*	
		0.00	0.00	-125.71	1.83	998.00	0.88	-180.86	2.63	1106.41	0.98	*N.E.*	
Oakland, MI	424	-0.50	-0.26	*N.E.*		*N.E.*		*N.E.*		0.00	0.00	*N.E.*	
		-363.64	-0.55	*N.E.*		*N.E.*		*N.E.*		0.00	0.00	*N.E.*	
Polk, IA	19281	-1.25	-1.01	0.41	0.49	-2.01	-1.62	0.06	0.07	0.19	0.16	0.00	0.00
		-36650.07	-3.26	-5408.38	-1.85	-7264.26	-0.65	201.21	0.07	5284.44	0.47	0.00	0.00
Pottawattamie, IA	4323	0.74	0.29	0.01	0.36	1.68	0.67	-0.05	-1.71	3.89	1.54	-0.05	-1.71
		-6163.33	-3.87	-7.48	-0.11	1394.84	0.88	26.36	0.40	1913.53	1.20	26.36	0.40
Ramsey, MN	10101	-0.28	-0.02	2.93	0.14	-0.16	-0.01	0.00	0.00	-0.12	-0.01	0.00	0.00
		-44010.86	-0.54	-134070.83	-2.76	-7587.52	-0.09	0.00	0.00	-7281.41	-0.09	0.00	0.00
Saginaw, MI	19084	0.13	0.09	*N.E.*		*N.E.*		*N.E.*		*N.E.*		*N.E.*	
		-410.25	-0.02	*N.E.*		*N.E.*		*N.E.*		*N.E.*		*N.E.*	
Scott, MN	11704	0.43	0.05	1.06	0.17	1.54	0.19	-2.18	-0.34	0.49	0.06	-2.18	-0.34
		-58468.63	-1.71	-1549.60	-0.03	-7198.97	-0.21	46704.56	0.77	-13601.79	-0.40	46704.56	0.77
St. Louis, MO	29338	0.17	0.07	*N.E.*		*N.E.*		*N.E.*		*N.E.*		*N.E.*	
		-2637.90	-0.04	*N.E.*		*N.E.*		*N.E.*		*N.E.*		*N.E.*	
Story, IA	11850	-0.33	-0.30	0.37	0.49	-0.60	-0.53	-0.22	-0.29	0.00	0.00	0.01	0.01
		-14381.63	-4.79	-2583.67	-1.28	-1597.77	-0.53	-583.37	-0.29	-9.08	0.00	23.30	0.01
Tuscola, MI	2823	0.28	0.12	*N.E.*		*N.E.*		*N.E.*		0.00	0.00	*N.E.*	
		-128.30	-0.02	*N.E.*		*N.E.*		*N.E.*		0.00	0.00	*N.E.*	
Washington, MN	3283	-0.18	-0.02	1.79	0.41	4.24	0.44	0.53	0.12	-0.20	-0.02	0.02	0.00
		-9547.82	-0.42	-7099.00	-4.83	8530.36	0.38	177.75	0.12	-3080.53	-0.14	6.92	0.00
Wayne, MI	89	0.00	0.00	*N.E.*		*N.E.*		*N.E.*		-2.44	-0.42	*N.E.*	
		0.00	0.00	*N.E.*		*N.E.*		*N.E.*		96.77	0.19	*N.E.*	
Woodbury, IA	6789	-0.16	-0.19	0.18	0.56	0.00	0.00	0.09	0.28	0.21	0.24	0.00	0.00
		-5253.64	-2.79	-707.23	-2.66	541.93	0.29	73.25	0.28	859.85	0.46	-0.08	0.00